DOCUMENT RESUME

ED 205 575 TM 810 462

AUTHOR	Samejima, Fumiko: Trestman, Robert L.
TITLE	Analysis of Iowa Data I: Initial Study and Findings. Research Report 80-1.
INSTITUTION	Tennessee Univ., Knoxville. Dept. of Psychology.
SPONS AGENCY	Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
PUB DATE	Apr 80
CONTRACT	N00014-77-C-0360
NOTE	154p.
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	*Achievement Tests: *Data Analysis: *Goodness of Fit: Junior High Schools: *Latent Trait Theory: Mathematical Models: Multiple Choice Tests
IDENTIFIERS	Chi Square Test: *Iowa Tests of Basic Skills: *Three Parameter Model

ABSTRACT

The first step of the data analysis with respect to the eventual application of the various new methods in latent trait theory is here initiated. The data are a set of approximately 500 item responses of each of 7,439 examinees to the Iowa Tests of Basic Skills, Form 6, on one of three difficulty levels, which correspond to the ages of 11, 12 and 13. The eleven subtests of this battery cover such areas as vocabulary, reading comprehension, language, mathematics and work-study skills, and have multiple choice formats, with four alternatives, except for one subtest in which there are five. After the initial tabulation, which includes various frequency distributions and the evaluation of its results, the elimination of certain examinees, who obviously skipped excessive numbers of items, is decided. An important question is whether the examinees' behavior fits the Knowledge and Random Guessing Principle, and then the three-parameter normal ogive or logistic model. The chi-square goodness of fit test provides surprisingly little support. Index k*, which was introduced on an earlier stag of this research, was also used for this purpose. (Author/BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

RESEARCH REPORT 80-1

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED205575

ANALYSIS OF IOWA DATA I: INITIAL STUDY AND FINDINGS

FUMIKO SAMEJIMA

AND

ROBERT L. TRESTMAN

DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF TENNESSEE
KNOXVILLE, TENN. 37916

APRIL, 1980

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

the Office of
Naval Research

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 462

Prepared under the contract number N00014-77-C-0360,
NR 150-402 with the
Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research

Approved for public release; distribution unlimited.
Reproduction in whole or in part is permitted for
any purpose of the United States Government.

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Research Report 80-1	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (*and Subtitle*) Analysis of Iowa Data I: Initial Study and Findings		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dr. Fumiko Samejima and Robert L. Trestman		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0360
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Tennessee Knoxville, TN 37916		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS PE: 61153N; PROJ: RR 042-04 TA: RR 042-04-01 WU: NR 150-402
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research (Code 458) Arlington, VA 22217		12. REPORT DATE 15 April, 1980
		13. NUMBER OF PAGES 147
14. MONITORING AGENCY NAME & ADDRESS(*if different from Controlling Office*)		15. SECURITY CLASS. (*of this report*) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (*of this Report*)

Approved for public release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States government.

17. DISTRIBUTION STATEMENT (*of the abstract entered in Block 20, if different from Report*)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (*Continue on reverse side if necessary and identify by block number*)

Operating Characteristic Estimation
Tailored Testing
Latent Trait Theory

20. ABSTRACT (*Continue on reverse side if necessary and identify by block number*)

(Please see reverse side)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S N 0102-LF-014-6601

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

The first step of the analysis of Iowa Data with respect to the eventual application of the various new methods in latent trait theory is here initiated. The data are a set of approximately 500 item responses of each of 7,439 examinees to the Iowa Tests of Basic Skills, Form 6, on one of three difficulty levels, which correspond to the ages of 11, 12 and 13. The eleven subtests of this battery cover such areas as vocabulary, reading comprehension, language, mathematics and work-study skills, and have multiple choice formats, with 4 alternatives, except for one subtest in which there are five. After the initial tabulation, which includes various frequency distributions and the evaluation of its results, the elimination of certain examinees, who obviously skipped excessive numbers of items, is decided. An important question is whether the examinees' behavior fits the Knowledge and Random Guessing Principle, and then the three-parameter normal ogive or logistic model. The Chi-Square Goodness of Fit Test provides surprisingly little support. Index k*, which was introduced on an earlier stage of this research, was also used for this purpose.

ANALYSIS OF IOWA DATA I: INITIAL STUDY AND FINDINGS

ABSTRACT

The first step of the analysis of Iowa Data with respect to the eventual application of the various new methods in latent trait theory is here initiated. The data are a set of approximately 500 item responses of each of 7,439 examinees to the Iowa Tests of Basic Skills, Form 6, on one of three difficulty levels, which correspond to the ages of 11, 12 and 13. The eleven subtests of this battery cover such areas as vocabulary, reading comprehension, language, mathematics and work-study skills, and have multiple choice formats, with 4 alternatives, except for one subtest in which there are five. After the initial tabulation, which includes various frequency distributions and the evaluation of its results, the elimination of certain examinees, who obviously skipped excessive numbers of items, is decided. An important question is whether the examinees' behavior fits the Knowledge and Random Guessing Principle, and then the three-parameter normal ogive or logistic model. The Chi-Square Goodness of Fit Test provides surprisingly little support. Index k*, which was introduced on an earlier stage of this research, was also used for this purpose.

The research was conducted at the principal investigator's laboratory, 409 Austin Peay Hall, Department of Psychology, University of Tennessee, Knoxville, Tennessee. Those who worked in the laboratory and helped the authors in various ways for this research include Paul S. Changas, Pamela Welch, Date Furlan, C. I. Bonnie Chen, Philip Livingston, Linda Lance, and Melanie Perkins.

ACKNOWLEDGEMENT

The authors are obliged to Professor William Coffman of the University of Iowa and Director of the Iowa Testing Programs for providing us with valuable data for our research.

A special word of appreciation is due to Dr. Charles Davis of the Office of Naval Research, who arranged the initial contact between the principal investigator and Professor William Coffman.

The work was divided between the two authors in such a way that the senior author designed the research, introduced new concepts such as Informative Distractor Model and Equivalent Distractor Model, and so forth, and the junior author engaged in the actual work of data analysis. The bibliography on the three-parameter logistic model in Appendix I was made by the junior author.

TABLE OF CONTENTS

Page

I Introduction 1

II Tests 5

III Data and Their Basic Tabulations 12

IV Informative Distractor Model or Equivalent Distractor Model? 36

V Comparisons of the Results of Test Items Which Were Administered to Levels 11, 12, and 13 91

VI Discussion and Conclusions 114

References 115

Appendix I 117

Appendix II 123

Appendix III 127

Appendix IV 135

I Introduction

There have been many research projects in the history of psychology, in which researchers attempt to measure mental abilities by collecting data through the administration of a battery of tests. Even before the American Psychological Association adopted the concepts of reliability and validity coefficients in 1954, classical test theory, with all its variations, has practically the sole theoretical foundation for those research projects, and for a long time it has been the theory of mental measurement.

There are many deficiencies and limitations in classical mental test theory, however. The weakest points in classical mental test theory may be: 1) it heavily depends upon a specific population, or group, of examinees, and defines the characteristics of a test, or a test item, in relation with a specific group of examinees by naming them as if they were sole properties of a test, or a test item, while in reality they are the products of the interaction between the test, or the test item, and the group of examinees; and 2) it uses a single measure, the correlation coefficient, for its important concepts such as the reliabilty and validity of a test and treats it as if it were a magic number, while the fact is that many complexities are ignored and distorted because of its use. Thus many researchers have misled and lost themselves in the cobweb of classical mental test theory without being aware of the fact that they have produced nothing but artifacts (cf. Samejima, 1977a).

Modern mental test theory, or latent trait theory, on the other hand, is based upon sounder axioms and rationale, and has gradually been invading the realm of classical mental test theory in the past decade, in the area of applied mental measurement. A relatively few, more sophisticated researchers have started conducting their research projects on mental measurement which is based upon latent trait theory. Among them, Rasch model (Rasch, 1960) may be the most popular model, because of its mathematical simplicity and easiness. When they deal with data in which multiple-choice test items are used, however, most researchers have turned to the three-parameter normal ogive, or logistic, model (Birnbaum, 1968), which is based upon the knowledge or random guessing principle in the context of latent trait theory. It is assumed that the examinee either knows the answer to a given multiple-choice test item, or guesses randomly, in selecting one of the given alternative answers. Thus the conditional probability, given ability, with which the examinee answers the item correctly is greater than the one with which the examinee knows the answer. When we consider this conditional probabilty as a function of ability, it is called the item characteristic function (Lord and Novick, 1968). In the three-parameter normal ogive, or logistic, model, this item characteristic function for a multiple-choice item is strictly increasing in ability, but its slope is less than the one for the item characteristic function in the normal ogive, or logistic, model for the free-response test item. A bibliography of applied research based upon the three-parameter models is presented in Appendix I , together with one for theoretical works concerning the three-parameter models, which consists of papers selected, mainly, from recent issues of leading journals.

In spite of the popularity of the three-parameter logistic model, except for Lord (Lord, 1970), the researchers rarely have tried to validate the model in relation to their own data; they simply adopted the model, and assumed its validity. Such attitudes of researchers cannot be considered as being scientific, and it is imperative that more people should turn their attention to model validation, or to the search for a suitable model, or models, before they adopt one for their research. It was one of her considerations when the principal investigator started the series of research on theory and method of estimating the operating characteristics without assuming any mathematical form several years ago (Samejima, 1977a, 1977b, 1978a, 1978b, 1978c, 1978d, 1978e, 1978f). She has also proposed a new family of models for the multiple-choice test items (Samejima, 1979, 1980), in which she treats the multiple-choice item as something more than a blurred image of the free-response test item, which the three-parameter normal ogive, or logistic, model presumes. Thus in these new models each alternative wrong answer is considered to be a valuable information source, in addition to the correct answer.

In the present study of mental measurement based upon the analysis of the Iowa Test data, unlike most research projects, we shall avoid setting any prior mathematical models, and start with the careful examination of our data, in order to determine which direction of research we should take. In other words, we are intentionally taking a slow process, and by no means are we anxious to extract hasty conclusions. This may seem to be tedious, but for truly scientific purposes of research this is the fastest way, if we wish to conduct research without sacrificing our conscience and with the promise of fruitfulness.

We hypothesize two main directions as our choices. One is the direction which leads to Informative Distractor Model, and the other is the direction which leads to Equivalent Distractor Model. Note, however, that they are not specific mathematical models, but a very general categorization of ideas concerning the behavior of the multiple-choice test item. If, for instance, evidence clearly indicates the former direction, then we shall follow that direction to investigate the behavior of each test item more specifically. If the latter direction proved to be true, then we shall head for that direction and investigate the specifics about each test item. If it turns out that no single, general direction is indicated, then we shall depend upon the strategy of adopting theory and method which enable us to deal with test items following either general direction.

The present research is only the beginning of the analysis of the Iowa Test data, and more research is needed to add to it in the near future.

II Tests

The battery of tests used here is the Iowa Tests of Basic Skills, Form 6, Levels 9-14. These tests have been designed, constructed, and revised at the College of Education of the University of Iowa since 1935, with the general school population in mind, and for students of ages nine through fourteen, or grades three through nine. All technical information in this paper has been taken from either Form 6 itself (Hieronymous and Lindquist, 1971), or its Teacher's Manual (Iowa Basic Skills Testing Program, 1971).

There are eleven tests in the battery, each of which focuses on a different basic skill. For convenience, hereafter, we shall call these separate tests subtests, in order to avoid the confusion which might occur when we refer to both the total test battery and each test in the battery. Following the Teacher's Manual, the descriptions and abbreviations of these eleven subtests, together with their administration schedule and working times, are tabulated and presented in Table 2-1. All the test items are power test items with multiple-choice format, with five alternative answers for the items in Subtest L1, and with four alternatives for those in the other ten subtests. Within each subtest, test items are arranged in the ascending order of difficulty. These eleven subtests are designed to cover all major areas of academic interest for the grades three through nine. The separate directions for these eleven subtests are presented in Appendix II.

TABLE 2-1

Administration Sessions, Time Limits and Subtests of Iowa Tests of Basic Skills.

Administration Session	Working Time (Minutes)	Subtest
First Session 85 Minutes	17 55	V: Vocabulary R: Reading Comprehension
Second Session 80 Minutes	12 15 20 20	L-1: Spelling L-2: Capitalization L-3: Punctuation L-4: Usage
Third Session 85 Minutes	30 20 30	W-1: Map Reading W-2: Reading Graphs and Tables W-3: Knowledge and Use of Reference Materials
Fourth Session 65 Minutes	30 30	M-1: Mathematics Concepts M-2: Mathematics Problem Solving

The numbers of test items contained by the eleven separate subtests are 114, 178, 114, 102, 102, 86, 89, 74, 141, 136 and 96, respectively, following the order of subtests given in Table 2-1. They are all presented in a single test booklet of ninety-six pages. For each of the five levels, 9 through 14, only a subset of each subtest is administered. For example, for Subtest V, items 1 through 31 are given as Level 9, items 11 through 48 are administered as Level 10, items 24 through 66 are given as Level 11, and so forth. Different answer forms for the computerized scoring are provided for the six different levels. The standardized administration schedule and the working time for each subtest are presented in Table 2-1. For the entire test battery, the time required for the administration of each level of test is four hours and thirty-nine minutes. It is recommended that the test be administered on four consecutive days. It may occasionally have been presented on four consecutive half-days, but never within one day.

There are three options available to the classroom teachers in the administration of the test. Among them, graded testing is the most common one, in which a single level of test is administered to all the students in a given classroom. In this situation, Level 9 is given to the third graders, Level 10 to the fourth graders, Level 11 to the fifth graders, Level 12 to the sixth graders, Level 13 to the seventh graders, and Level 14 to both the eighth and ninth graders. The second option is out-of-level testing. In this situation, the teacher selects an optimal level for his or her class, and administers that level of test, regardless of the examinees' nominal grade. The third option is individualized testing, in which each student in the class is given the level of test which is most suitable for his or her level of development. We notice that, even in this third situation, it is possible to administer the test as a group test with a single supervisor, since the administration schedule is standardized for all levels, as we can see in Table 2-1.

In our data, only the tests of Levels 11, 12 and 13 were used. In most cases, the first form of administration, i.e., graded testing, was adopted. There are a few exceptions, however, in which the third form was used. The numbers of test items contained in these three levels of test are 461, 487 and 500, respectively. Table 2-2 presents the number of test items in each of the eleven subtests, for each of the three levels. A graphical representation is made in Figure 2-1, to show how these three subsets of test items in each subtest overlap among the three levels. The numbers of test items in each subtest, which are included in all the three levels, in two adjacent levels, and in single levels, are shown in Table 2-3.

TABLE 2-2

Number of Items in Each of the Eleven Subtests and in Total for Each of the Three Levels, 11, 12 and 13.

Subtest Level	V	R	L1	L2	L3	L4	W1	W2	W3	M1	M2	Total
11	43	74	43	40	40	32	36	26	56	42	29	461
12	46	76	46	42	42	32	40	28	59	45	31	487
13	48	78	48	43	43	32	41	28	59	48	32	500

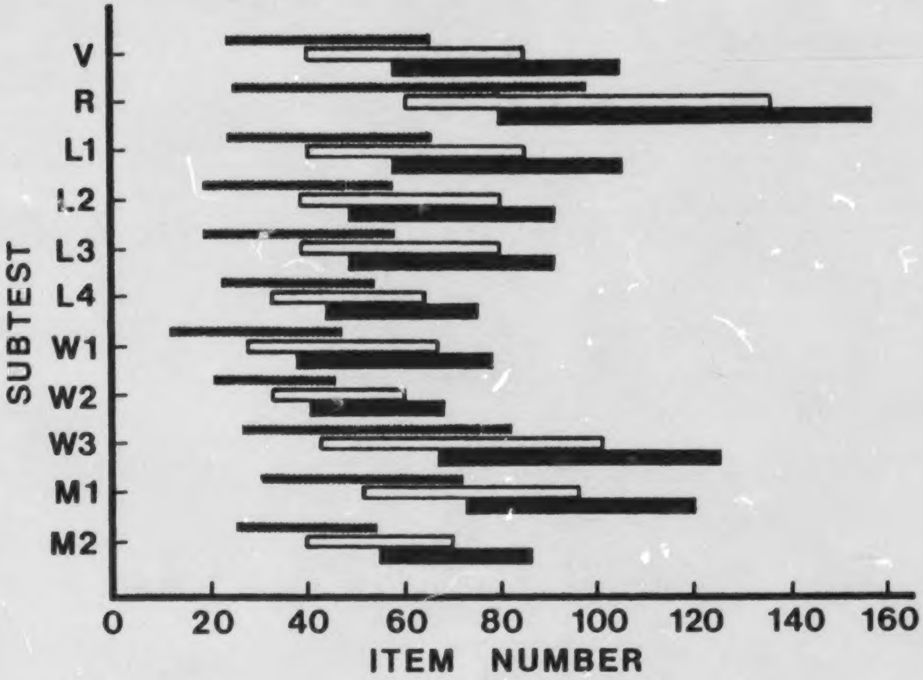


FIGURE 2-1

Test Items of Each of the Eleven Subtests Administered to Each of Levels 11, 12 and 13, Which Are Represented by Shaded, Hollow, and Solid Bars, Respectively.

TABLE 2-3

Frequency Distribution of Items Which Were Administered to More Than One Level, As Well As Those Which Were Administered to One Level Only.

Level Subtest	11,12,13	11,12	12,13	12 Only	13 Only	Total
V	9	27	28	16	20	100
R	19	38	57	36	21	171
L1	9	27	28	16	20	100
L2	10	20	32	20	11	93
L3	10	20	32	20	11	93
L4	11	22	21	10	11	75
W1	10	20	30	16	11	87
W2	6	14	20	12	8	60
W3	16	40	35	16	24	131
M1	0	21	24	21	24	90
M2	0	15	16	14	16	61
Total	100	264	323	197	177	1,061

III Data and Their Basic Tabulations

Data were collected in three different school systems in the State of Iowa, in the years 1971 through 1977. In their original form, the total number of examinees, including both boys and girls, is 7,581. Out of these people, 28 students took Level 9 Test and 114 took Level 10 Test. Since these are relatively small numbers, we decided to exclude them from our original group of examinees. The other 7,439 examinees are classified into three subgroups, i.e., 2,460 students who took Level 11 Test, 2,452 who took Level 12 Test, and 2,527 who took Level 13 Test. Hereafter, we shall call observations concerning these 7,439 examinees the original data.

Table 3-1 presents the frequency distribution of the test items for each of the eleven subtests with respect to the percentage of examinees who answered correctly, for each of Levels 11, 12 and 13. We can see that the configurations of these eleven frequency distributions are alike across the three levels, and, except for Subtest L1 for Level 13 and Subtest M2 for Levels 12 and 13, the medians of these frequency distributions are somewhat higher than 50 percent.

In these frequency distributions, all the no responses were treated as incorrect answers. Table 3-2 presents the frequency distributions of test items for each of the eleven subtests with respect to the percentage of examinees who answered in one way or another, for each of the three levels. We notice that, while

TABLE 3-1

Frequency Distribution of Items for Each of the Eleven Subtests with Respect to the Percentage of Examinees Answering Correctly. Each Interval of Percentage Is Greater than or Equal to the Lower End and Less than the Upper End. (Original Data)

Level 11

	Percentage	Subtest											Total
		V	R	L1	L2	L3	L4	W1	W2	W3	M1	M2	
1	0.0 - 5.0												0
2	5.0 - 10.0												0
3	10.0 - 15.0												0
4	15.0 - 20.0		1	2	1								4
5	20.0 - 25.0		1	1	1		1	1			1	3	9
6	25.0 - 30.0	2	2				1	2			1	2	10
7	30.0 - 35.0		4	3			1	1		1	2	3	15
8	35.0 - 40.0	3	2	1	1	3	2	2		2	4	2	22
9	40.0 - 45.0	5	5	11	4	5	5	1	3	3	3	2	47
10	45.0 - 50.0	4	12	3	5	3	5	2	3	8	4	1	50
11	50.0 - 55.0	3	6	3	3	2	4	4	8	9	8	2	52
12	55.0 - 60.0	11	5	4	4	4	4	5	2	13	2	3	57
13	60.0 - 65.0	4	5	2	9	5	5	7	3	7	2	3	52
14	65.0 - 70.0	6	5	2	5	10	3	3	1	9	5	3	52
15	70.0 - 75.0	2	9	6	2	1	1	2	1	3	4	3	34
16	75.0 - 80.0	2	6	5	3	6		1	1	1	3	2	30
17	80.0 - 85.0	1	7		1	1		2	2		1		15
18	85.0 - 90.0		2		1			1	1		1		6
19	90.0 - 95.0		2					2	1		1		6
20	95.0 -100.0												0
	Total	43	74	43	40	40	32	36	26	56	42	29	461

TABLE 3-1 (Continued): Level 12.

	Percentage	V	R	L1	L2	L3	L4	W1	W2	W3	M1	M2	Total
1	0.0 - 5.0												0
2	5.0 - 10.0												0
3	10.0 - 15.0												0
4	15.0 - 20.0					1							1
5	20.0 - 25.0	1	1	1	1	1					2	1	8
6	25.0 - 30.0		1	5	1		3	1		1	2	2	16
7	30.0 - 35.0	2	4	6			1	5	2	3	1	1	25
8	35.0 - 40.0	6	5	5	2	3	2	1		4	2	4	34
9	40.0 - 45.0		3	2	3	1	1	3	4	5	5	3	30
10	45.0 - 50.0	5	9	3	1	6	6	7	1	5	3	5	51
11	50.0 - 55.0	2	14	4	2	8	5	5	1	1	8	3	53
12	55.0 - 60.0	7	8	8	6	6	5	3	4	6	4	3	60
13	60.0 - 65.0	8	12	[illegible]	6	5	5	2	6	1	3	1	52
14	65.0 - 70.0	4	4	4	5	4	3	1	3	11	4		43
15	70.0 - 75.0	5	7	1	4	3	1	8	3	7	3	5	47
16	75.0 - 80.0	4	3	2	8	3		3	2	7	4	3	39
17	80.0 - 85.0	2	2	2	2	1			1	6	2		18
18	85.0 - 90.0		2					1	1	2	2		8
19	90.0 - 95.0		1		1								2
20	95.0 -100.0												0
	Total	46	76	46	42	42	32	40	28	59	45	31	487

TABLE 3-1 (Continued): Level 13.

	Percentage	Subtest V	 R	 L1	 L2	 L3	 L4	 W1	 W2	 W3	 M1	 M2	Total
1	0.0 - 5.0												0
2	5.0 - 10.0												0
3	10.0 - 15.0		1										1
4	15.0 - 20.0					1	1		1				3
5	20.0 - 25.0			3		2		1	1	1		4	12
6	25.0 - 30.0	1	2	3	1	2	2	1		1		5	18
7	30.0 - 35.0	4	3	5	2	1	5	2	2	1	7	2	34
8	35.0 - 40.0	3	6	5	2	3	2	5	3	6	3		38
9	40.0 - 45.0	4	9	11	2	4	3	2	2	8	7	3	55
10	45.0 - 50.0	4	4	3	4	3	1	3	3	3	6	5	39
11	50.0 - 55.0	3	6	4	5	7	5	6	2	10	2	5	55
12	55.0 - 60.0	10	11	2	5	7	4	6		7	5	3	60
13	60.0 - 65.0	2	14	4	3	6	5	3	2	4	5	1	49
14	65.0 - 70.0	5	8	4	5	3	2	3	4	2	4		40
15	70.0 - 75.0	7	8	3	5	2	2	3	2	8	6	1	47
16	75.0 - 80.0	1	3	1	5			2	3	3	2		20
17	80.0 - 85.0	2	2		3	2		3	2	2		2	18
18	85.0 - 90.0	2	1		1				1	3	1	1	10
19	90.0 - 95.0							1					1
20	95.0 -100.0												0
	Total	48	78	48	43	43	32	41	28	59	48	32	500

TABLE 3-2

Frequency Distribution of Items for Each of the Eleven Subtests with Respect to the Percentage of Examinees Answering, Though Not Necessarily Correctly. Each Interval of Percentage Is Greater than the Lower End and Less than or Equal to the Upper End. (Original Data)

Level 11

Percentage / Subtest	0-90%	90-91%	91-92%	92-93%	93-94%	94-95%	95-96%	96-97%	97-98%	98-99%	99-100%	99.0-99.5%	99.5-100.0%
V	1	1	3		2	2	2	2	2	8	20	5	15
R	4	2	2	1	4	3	3	1	7	7	40	7	33
L1	5	1	1	1	1	2	1	2	3	5	21	9	12
L2			1	2	1	1	3	2	5	4	21	5	16
L3								4	5	7	24	9	15
L4									3	6	23	15	8
W1	7		1		1	1	1	1	3	5	16	7	9
W2				1	2	1	1	1	2	4	14	3	11
W3	14	2		2	1	2	1	3	2	10	19	15	4
M1		1	2		2	1	2	5	1	7	21	8	13
M2	3	1		1	1	1	2	2	1	5	12	6	6
Total	34	8	10	8	15	14	16	23	34	68	231	89	142

TABLE 3-2 (Continued): Level 12.

Percentage / Subtest	0-90%	90-91%	91-92%	92-93%	93-94%	94-95%	95-96%	96-97%	97-98%	98-99%	99-100%	99.0-99.5%	99.5-100.0%
V					1	3	2	2	2	5	31	5	26
R		4	2	3	4	2	1	5	5	5	45	8	37
L1	6		1	1	1	1	2	2	2	4	26	3	23
L2						2	3	3	2	4	28	3	25
L3										8	34	3	31
L4										1	31	7	24
W1	8		1		1	1	2	3	1	2	21	5	16
W2		1	1	1		1		2	2	2	18	2	16
W3			2	2	2	3	1	2	4	5	38	7	31
M1					2	2	1	3	5	4	28	11	17
M2	1		1	1	1	1	1	1	2	3	19	11	8
Total	15	5	8	8	12	16	13	23	25	43	319	65	254

TABLE 3-2 (Continued): Level 13.

Percentage Subtest	0-90%	90-91%	91-92%	92-93%	93-94%	94-95%	95-96%	96-97%	97-98%	98-99%	99-100%	99.0-99.5%	99.5-100.0%
V								4	6	11	27	27	
R						7	4	6	4	21	36	36	
L1	1	3		2	2	2	1	3	4	5	25	25	
L2								5	5	6	27	27	
L3										9	34	34	
L4										4	28	28	
W1	5	1	2	1		2		2	4	6	18	18	
W2						2	1	2	2	4	17	17	
W3				1	3	3	3	3	4	11	31	31	
M1						3	2	3	2	9	29	29	
M2			1	2	1	1		1	3	9	14	14	
Total	6	4	3	6	6	20	11	29	34	95	286	286	0

the majority of items, i.e., 231 out of 461 for Level 11, 319 out of 487 for Level 12, and 286 out of 500 for Level 13, have attracted some answers from 99 percent or more of the examinees, there are some other items, i.e., 34 for Level 11, 15 for Level 12, and 6 for Level 13, to which only less than 90 percent of the examinees have responded. The cumulative frequency of test items to which one of the six given percentages, or more, of the examinees responded is presented in Table 3-3, for each of the eleven subtests, and for each of the three levels.

The breakdown of the last column of Table 3-2 into two subcategories, i.e., 99.0 to 99.5 percent and 99.5 to 100.00 percent, which are added to its right as two additional columns of Table 3-2, has revealed that, for Level 13, there are no items to which 99.5 percent, or more, of the examinees responded. This result looks rather strange, and has urged us to shift our attention back to our raw data. As the result, it was found out that there are a small number of examinees who did not respond to a substantially large number of test items. Table 3-4 presents the frequency distributions of examinees with respect to the number of unanswered test items for Levels 11, 12 and 13. We can see in this table that, while as many as 7,010 examinees out of 7,439 left only 49 or less test items unanswered, there also are 162 examinees who did not respond to as many as 100, or more, test items. Our raw data show there are some examinees included who skipped an entire subtest, or more than one entire subtest.

TABLE 3-3

Frequency Distribution of Items Which Were Answered, Though Not Necessarily Correctly, by Specified Percentages of Examinees or More. (Original Data)

Level 11

Subtest	Number of Items	Minimum Percentage Response					
		90%	92%	94%	95%	96%	97%
V	43	42	38	36	34	32	30
R	74	70	66	61	58	55	54
L1	43	38	36	34	32	31	29
L2	40	40	39	36	35	32	30
L3	40	40	40	40	40	40	36
L4	32	32	32	32	32	32	32
W1	36	29	28	27	26	25	24
W2	26	26	26	23	22	21	20
W3	56	42	40	37	35	34	31
M1	42	42	39	37	36	34	29
M2	29	26	25	23	22	20	18
Total	461	427	409	386	372	356	333

TABLE 3-3 (Continued): Level 12.

Subtest	Number of Items	Minimum Percentage Response 90%	92%	94%	95%	96%	97%
V	46	46	46	45	42	40	38
R	76	76	70	63	61	60	55
L1	46	40	39	37	36	34	32
L2	42	42	42	42	40	37	34
L3	42	42	42	42	42	42	42
L4	32	32	32	32	32	32	32
W1	40	32	31	30	29	27	24
W2	28	28	26	25	24	24	22
W3	59	59	57	53	50	49	47
M1	45	45	45	43	41	40	37
M2	31	30	29	27	26	25	24
Total	487	472	459	439	423	410	387

TABLE 3-3 (Continued): Level 13.

Subtest	Number of Items	Minimum Percentage Response 90%	92%	94%	95%	96%	97%
V	48	48	48	48	48	48	44
R	78	78	78	78	71	67	61
L1	48	47	44	40	38	37	34
L2	43	43	43	43	43	43	38
L3	43	43	43	43	43	43	42
L4	32	32	32	32	32	32	32
W1	41	36	33	32	30	30	28
W2	28	28	28	28	26	25	23
W3	59	59	59	55	52	49	46
M1	48	48	48	48	45	43	40
M2	32	32	31	28	27	27	26
Total	500	494	487	475	455	444	415

TABLE 3-4

Frequency Distribution of Examinees with Respect to the Number of Items Which Were Left Unanswered for Each of the Three Levels, 11, 12, and 13. (Original Data)

Number Unanswered	Level 11	Level 12	Level 13	Total
0-49	2271	2326	2413	7010
50-99	120	85	62	267
100-199	55	34	29	118
200-299	13	7	18	38
300-500	1	0	5	6
Total	2460	2452	2527	7439

Table 3-5 presents the frequency distribution of such examinees with respect to the number of subtests which were omitted entirely, for each of the three levels. We can see in this table that one examinee who did not participate in the testing at all is included in our group of examinees. The relationship between the omission of one or more entire subtests and the number of unanswered items is shown in Table 3-6 in the form of the frequency distribution of these 67 examinees with respect to the level and the number of unanswered test items. Comparison of this table with Table 3-4 indicates that the six examinees who left 300 or more test items unanswered are among those who omitted, at least, one entire subtest, and as many as 27 out of 38 examinees who omitted 200 to 299 test items also belong to this subgroup of examinees. A frequency distribution similar to Table 3-6 was made for the 126 examinees who left, at least, one half of a subtest, but not the entire subtest, unanswered, and is presented as Table 3-7. This additional information indicates that, if we exclude all the examinees who left, at least, one half of a subtest unanswered from our total group of examinees, then the number of examinees who left 200 or more test items unanswered will become zero, and only 28 examinees, who omitted more than 100, but less than 200, test items, will be included. For this reason, we have decided to exclude these 193 examinees from our original group of examinees for further analysis. Hereafter, we shall call observations concerning the remaining 7,246 examinees the revised data, to

TABLE 3-5

Frequency Distribution of Examinees Having Omitted, at Least, One Entire Subtest with Respect to the Total Number of Unanswered Subtests for Each of the Three Levels, 11, 12, and 13. (Original Data)

Level	Number of Subtests Omitted Entirely											Total
	1	2	3	4	5	6	7	8	9	10	11	
11	1	3	1	3	1	2	3				1	15
12	2	5		1	2							10
13	3	10	7	2	8	10		1	1			42
Total	6	18	8	6	11	12	3	1	1	0	1	67

TABLE 3-6

Frequency Distribution of Examinees Who Omitted, at Least, One Entire Subtest with Respect to the Total Number of Items Omitted, for Each of the Three Levels, 11, 12, and 13. (Original Data)

Number Unanswered	Level 11	Level 12	Level 13	Total
0-49	0	0	0	0
50-99	2	5	5	12
100-199	6	2	14	22
200-299	6	3	18	27
300-500	1	0	5	6
Total	15	10	42	67

TABLE 3-7

Frequency Distribution of the Examinees Having Left, at Least, One Half of a Subtest Unanswered, But Not Omitted Any Entire Subtest, with Respect to the Number of Unanswered Items, for Each of the Three Levels, 11, 12, and 13. (Original Data)

Number Unanswered	Level 11	12	13	Total
0-49	9	1	3	13
50-99	25	5	4	34
100-199	40	19	9	68
200-299	7	4	0	11
300-500	0	0	0	0
Total	81	29	16	126

distinguish themselves from the original data.

Table 3-8 presents the item identifications of the fifty-five test items, i.e., 34 for Level 11, 15 for Level 12, and 6 for Level 13, to which less than 90 percent of examinees responded in the original data, the percentages of examinees who answered in one way or another in the original data, and those in the revised data. We can see in this table that for most of these fifty-five test items the two percentages show a visible improvement caused by the exclusion of the 193 examinees. The frequency distributions of test items for the eleven subtests with respect to the percentage of examinees who answered in one way or another in the revised data are presented in Table 3-9, for each of the three levels. Comparison of this table with Table 3-2 reveals a substantial improvement in the percentage provided by the exclusion of the 193 examinees for all the three levels. Among others, we notice that the frequency of test items which were answered by 99 percent, or more, of examinees increased from 231 to 320 for Level 11, from 319 to 350 for Level 12, and from 286 to 377 for Level 13.

Table 3-10 presents the frequency distribution of test items for each of the eleven subtests with respect to the percentage of examinees who answered correctly, for each of Levels 11, 12 and 13, after the exclusion of the 193 examinees. There are a substantial number of categories whose frequencies changed from those in Table 3-1, although the two frequency distributions for each

TABLE 3-8

Fifty-five Test Items to Which Less Than Ninety Percent of Examinees Responded in One of the Three Levels in the Original Data, the Percentages of Responses in the Original Data, and Those in the Revised Data.

Item	Level 11		Level 12		Level 13	
	Original	Revised	Original	Revised	Original	Revised
V-66	89.1	91.4				
R-95	89.6	91.9				
R-96	89.2	91.5				
R-97	88.7	91.0				
R-98	88.3	90.7				
L1-62	89.4	91.8				
L1-63	88.3	90.7				
L1-64	87.5	90.0				
L1-65	86.3	88.7				
L1-66	84.8	87.3				
L1-80			89.7	90.7		
L1-81			88.9	89.9		
L1-82			87.9	88.9		
L1-83			87.0	88.0		
L1-84			86.2	87.2		
L1-85			85.4	86.4		
L1-105					89.7	90.6
W1-41	88.9	91.4				
W1-42	85.6	88.2				
W1-43	83.3	86.0				
W1-44	81.7	84.3				
W1-45	79.2	81.6				
W1-46	76.7	79.0				
W1-47	74.7	76.9				
W1-60			89.2	90.4		
W1-61			87.2	88.3		
W1-62			85.2	86.3		
W1-63			82.7	83.8		
W1-64			80.8	81.9		
W1-65			78.4	79.4		
W1-66			75.4	76.3		
W1-67			74.2	75.2		
W1-74					88.9	90.0
W1-75					87.3	88.4
W1-76					86.2	87.2
W1-77					85.0	86.1
W1-78					83.9	84.9
W3-69	90.0	92.3				
W3-70	89.3	91.5				
W3-71	88.8	91.0				
W3-72	87.9	90.2				
W3-73	87.1	89.4				
W3-74	86.8	89.0				
W3-75	86.0	88.3				
W3-76	84.9	87.2				
W3-77	84.0	86.3				
W3-78	83.5	85.8				
W3-79	82.8	85.0				
W3-80	82.3	84.5				
W3-81	81.6	83.8				
W3-82	81.1	83.3				
M2-52	87.9	90.0				
M2-53	85.9	87.9				
M2-54	83.7	85.7				
M2-69			89.4	90.3		

TABLE 3-9

Frequency Distribution of Items for Each of the Eleven Subtests with Respect to the Percentage of Examinees Answering, Though Not Necessarily Correctly. Each Interval of Percentage Is Greater than the Lower End and Less than or Equal to the Upper End.

Level 11

Percentage / Subtest	0-90%	90-91%	91-92%	92-93%	93-94%	94-95%	95-96%	96-97%	97-98%	98-99%	99-100%	Total
V			1	1	2	1	2	2	2	4	28	43
R		2	2	1	2	2	4	4	2	5	50	74
L1	3	1	1	1		1	2	2	1	4	27	43
L2					1	1	2	1	3	3	29	40
L3										5	35	40
L4										1	31	32
W1	6		1		1		1	1	1	2	23	36
W2						1	1	2	1	1	20	26
W3	10	1	2	2	1	1	2	1	2	3	31	56
M1				1	2		2	1	5	3	28	42
M2	3			1		1	1	1	2	2	18	29
Total	22	4	7	7	9	8	17	15	19	33	320	461

TABLE 3-9 (Continued): Level 12.

Percentage / Subtest	0-90%	90-91%	91-92%	92-93%	93-94%	94-95%	95-96%	96-97%	97-98%	98-99%	99-100%	Total
V						1	3	2	2	4	34	46
R			2	4	3	4	2	1	5	7	48	76
L1	5	1		1	1	1	1	2	2	3	29	46
L2							2	3	3	3	31	42
L3										4	38	42
L4											32	32
W1	7	1		1		1	1	2	2	2	23	40
W2			1	1	1		1		3	1	20	28
W3				1	2	3	3	1	2	4	43	59
M1						2	2	1	3	5	32	45
M2		1		2		1	1	1	2	3	20	31
Total	12	3	3	10	7	13	16	13	24	36	350	487

TABLE 3-9 (Continued): Level 13.

Percentage / Subtest	0-90%	90-91%	91-92%	92-93%	93-94%	94-95%	95-96%	96-97%	97-98%	98-99%	99-100%	Total
V									2	8	38	48
R							7	4	5	6	56	78
L1		1	3	1	1	1	3	1	3	4	30	48
L2									5	5	33	43
L3										1	42	43
L4											32	32
W1	5		1	1	2		2		2	4	24	41
W2							2	1	2	2	21	28
W3					1	3	3	3	3	4	42	59
M1							3	2	3	3	37	48
M2				1	2	1	1	1		4	22	32
Total	5	1	4	3	6	5	21	12	25	41	377	500

TABLE 3-10

Frequency Distribution of Items for Each of the Eleven Subtests with Respect to the Percentage of Examinees Answering Correctly. Each Interval of Percentage Is Greater than or Equal to the Lower End and Less than the Upper End.

Level 11

	Percentage	Subtest V	R	L1	L2	L3	L4	W1	W2	W3	M1	M2	Total
1	0.0 - 5.0												0
2	5.0 - 10.0												0
3	10.0 - 15.0												0
4	15.0 - 20.0		1	2	1								4
5	20.0 - 25.0		1	1	1		1	1			1	1	7
6	25.0 - 30.0	1	2				1	1				4	9
7	30.0 - 35.0	1	4	2			1	2		1	3	2	16
8	35.0 - 40.0	3	2	2	1	2	1	2		1	4	3	21
9	40.0 - 45.0	4	5	6	3	4	5	1	2	3	3	2	38
10	45.0 - 50.0	4	9	7	6	5	6	1	3	4	3		48
11	50.0 - 55.0	4	8	3	3		4	4	7	10	8	3	54
12	55.0 - 60.0	10	5	5	2	5	4	6	3	15	3	2	60
13	60.0 - 65.0	5	4	1	9	5	5	4	2	6	2	3	46
14	65.0 - 70.0	4	6	3	5	9	3	6	3	9	5	4	57
15	70.0 - 75.0	4	9	5	3	2	1	2	1	6	3	3	39
16	75.0 - 80.0	2	5	6	4	6		1	1	1	4	2	32
17	80.0 - 85.0	1	7		1	1		2	2		1		15
18	85.0 - 90.0		4		1	1		1	1		1		9
19	90.0 - 95.0		2					2	1		1		6
20	95.0 -100.0												0
	Total	43	74	43	40	40	32	36	26	56	42	29	461

TABLF 3-10 (Continued): Level 12.

	Percentage	Subtest											Total
		V	R	L1	L2	L3	L4	W1	W2	W3	M1	M2	
1	0.0 - 5.0												0
2	5.0 - 10.0												0
3	10.0 - 15.0												0
4	15.0 - 20.0					1							1
5	20.0 - 25.0	1	1	1	1	1					2	1	8
6	25.0 - 30.0		1	5	1		2	1		1	2	2	15
7	30.0 - 35.0	1	3	5			2	3	1	3	1	1	20
8	35.0 - 40.0	6	6	5	2	3	2	2	1	4	2	4	37
9	40.0 - 45.0	1	2	2	3	1	1	4	4	4	4	3	29
10	45.0 - 50.0	3	10	4	1	6	6	7	1	5	4	5	52
11	50.0 - 55.0	4	12	4	2	8	5	4	1	2	7	1	50
12	55.0 - 60.0	7	9	6	5	6	5	4	4	6	4	5	61
13	60.0 - 65.0	8	12	5	6	4	5	2	6	1	4	1	54
14	65.0 - 70.0	4	5	4	5	5	3	1	2	10	4		43
15	70.0 - 75.0	5	6	1	5	2	1	7	4	8	3	5	47
16	75.0 - 80.0	4	4	2	6	4		4	2	5	4	3	38
17	80.0 - 85.0	2	2	2	4				1	8	2		21
18	85.0 - 90.0		2			1		1	1	2	2		9
19	90.0 - 95.0		1		1								2
20	95.0 -100.0												0
	Total	46	76	46	42	42	32	40	28	59	45	31	487

TABLE 3-10 (Continued): Level 13.

	Percentage	V	R	L1	L2	L3	L4	W1	W2	W3	M1	M2	Total
		Subtest											
1	0.0 - 5.0												0
2	5.0 - 10.0												0
3	10.0 - 15.0		1										1
4	15.0 - 20.0					1	1		1				3
5	20.0 - 25.0			3		2		1	1	1		4	12
6	25.0 - 30.0	1	2	3	1	2	1	1		1		5	17
7	30.0 - 35.0	4	3	4	2	1	5	1	2	1	6	2	31
8	35.0 - 40.0	3	4	6	1	3	3	5	3	6	3		37
9	40.0 - 45.0	4	11	11	2	2	3	3	2	8	8	3	57
10	45.0 - 50.0	2	4	2	5	4	1	2	2	3	5	5	35
11	50.0 - 55.0	5	6	5	4	7	4	6	3	10	3	4	57
12	55.0 - 60.0	6	9	2	6	7	5	7		6	5	4	57
13	60.0 - 65.0	6	15	4	3	6	4	3	2	4	5	1	53
14	65.0 - 70.0	4	8	4	3	3	2	3	4	3	4		38
15	70.0 - 75.0	7	7	2	7	3	3	3	2	5	6	1	46
16	75.0 - 80.0	2	5	2	4			2	2	6	2		25
17	80.0 - 85.0	2	2		4	2		2	3	2		2	19
18	85.0 - 90.0	2	1		1			1	1	3	1	1	11
19	90.0 - 95.0							1					1
20	95.0 -100.0												0
	Total	48	78	48	43	43	32	41	28	59	48	32	500

subtest and each of the three levels are similar in configuration.

It should be noted that, even in the revised data, these percentages correct are not independent from the positions of the test items in each subtest. The relative frequencies of examinees who left the items unanswered are given in Appendix III, in which the test items are arranged in the order of presentation within each subtest, for each of the three levels. There is a distinct tendency that larger numbers of examinees did not respond to items which were presented later in each subtest. It is obvious, therefore, that, for these later items, the percentage for the correct answer is less than it should be in the ideally set free-response situation.

IV Informative Distractor Model or Equivalent Distractor Model?

By Informative Distractor Model, we mean the family of models in which we assume the existence of specific information obtainable from separate alternative answers, including the correct answer, of each multiple-choice test item. The family of models, including Models A, B and C, proposed by Samejima (Samejima, 1979) belongs to this general model.

In contrast to this, by Equivalent Distractor Model we mean the family of models in which no specific information is expected from separate incorrect answers, which are given as alternatives in the multiple-choice test item. Thus all the alternatives, except for the correct answer, of a given multiple-choice item are equivalent, since the information given by a specific alternative, or distractor, is not different from the one given by any one of the remaining wrong answers. The three-parameter normal ogive, or logistic, model belongs to this family of models. In this model, all the information provided by a given wrong answer is pure noise resulting from random guessing, and, therefore, the alternative is equivalent with any one of the remaining wrong answers. Note, however, that this type of model is not the only one included by the Equivalent Distractor Model. Suppose that the operating characteristic of each wrong answer of a given multiple-choice item includes some information about the examinee's ability, but all the operating characteristics, or plausibility curves, of the distractors are identical. In such a case, we can say that the item should belong to the Informative Distractor Model in the sense that these distractors provide us with some information concerning the examinee's ability. On the other hand, we can also say that the item should belong to the Equivalent Distractor Model, since each distractor does not have any specific information which distinguishes itself from the other distractors. For convenience, in the present paper, we shall take the second standpoint, defining the Informative Distractor Model in the narrower sense.

As an indicator which enables us to determine the direction each test item should take, we shall use Index k^*, which was proposed by Samejima (Samejima, 1980). This index was originally developed for the purpose of invalidating the knowledge or random guessing principle, upon which the three-parameter normal ogive, or logistic, model is based. It is also useful as the first step of searching the direction for each test item, with the two general models in mind.

Let $\bar{A}$ be the event that the examinee does not know the answer to the multiple-choice item g. We shall consider the probability space which consists of the subpopulation of such examinees. The conditional probability, $p(i|\bar{A})$, with which the examinee selects the alternative $i(=1,2,...,m_g$, or $m)$ of item g in this conditional probability space is given by

(4.1) $$p(i|\bar{A}) \begin{cases} = p_i[\sum_{i\neq R} p_i + p_R^*]^{-1} & i\neq R \\ = p_R^*[\sum_{i\neq R} p_i + p_R^*]^{-1} , & i=R \end{cases}$$

where p_i is the probability with which the examinee chooses the alternative i, and p_R^* is the probability with which the examinee guesses correctly for item g. Index k^* is defined in terms of these conditional probabilities, such that

(4.2) $$k^* = \exp[-\sum_{i=1}^{m} p(i|\bar{A})\cdot\log p(i|\bar{A})] = [\prod_{i=1}^{m} p(i|\bar{A})^{p(i|\bar{A})}]^{-1} .$$

It is obvious that this conditional probability, $p(i|\bar{A})$, for a wrong answer, or $i\neq R$, is proportional to p_i, since, according to the knowledge or random guessing principle, every examinee in the original population who has selected a wrong answer does not know the correct answer to item g, and, consequently, belongs to the subpopulation $\bar{A}$. On the other hand, we can write

(4.3) $$p_R^* \leq p_R ,$$

for examinees who have selected the correct answer R do not necessarily belong to the subpopulation $\bar{A}$.

Let θ be ability, or latent trait, and $P_g(\theta)$ be the item characteristic function of the multiple-choice item g. In the three-parameter normal ogive, or logistic, model, this item characteristic function is given by

$$P_g(\theta) = \Psi_g(\theta) + [1-\Psi_g(\theta)]c_g = c_g + [1-c_g]\Psi_g(\theta) \ , \tag{4.4}$$

where $\Psi_g(\theta)$ is the item characteristic function of item g when it is given as a free-response test item, which is specified by

$$\Psi_g(\theta) = (2\pi)^{-1/2} \int_{-\infty}^{a_g(\theta-b_g)} e^{-u^2/2}\, du \tag{4.5}$$

in the normal ogive model, and by

$$\Psi_g(\theta) = [1 + \exp\{-Da_g(\theta-b_g)\}]^{-1} \tag{4.6}$$

in the logistic model, and c_g is the guessing parameter which equals $1/m$. Let $f(\theta)$ denote the density function of ability θ , and p_R and p_g be the probabilities with which the examinee answers the item correctly in the multiple-choice situation and the free-response situation, respectively. Thus we can write

$$p_g = \int_{-\infty}^{\infty} \Psi_g(\theta)\, f(\theta)\, d\theta \ , \tag{4.7}$$

and

$$p_R = \int_{-\infty}^{\infty} P_g(\theta)\, f(\theta)\, d\theta = p_g + c_g(1-p_g) \ . \tag{4.8}$$

It is noted that the second term of the rightest hand side of (4.8) is the probability with which the examinee guesses correctly, i.e.,

p^*_R . We also notice that this quantity equals the probability with which the examinee chooses a distractor i $(\neq R)$. Thus we can write

$$p_i = c_g(1-p_g) = p^*_R \quad . \qquad (i \neq R) \tag{4.9}$$

It is obvious from (4.9), (4.1) and (4.2) that Index k* assumes its maximal value, m , when the three-parameter normal ogive, or logistic, model holds. Note, however, that it is a necessary condition for the validity of the model, but not a sufficient condition. Index k* can be used for the invalidation of the three-parameter normal ogive, or logistic, model, therefore, but not for the validation of the model, unless it is combined with other evidence.

In practice, we obtain the estimate of Index k* , by replacing $p(i|\overline{A})$ in (4.2) by its estimate, $\hat{p}(i|\overline{A})$. To obtain this estimate, we can use the frequency ratio, P_i , as the estimate of the probability, p_i , for each wrong answer, or distractor, i $(\neq R)$. We notice, however, that the estimate, P^*_R , for the probability p^*_R , is not directly observable from our data, and, in one way or another, must be defined indirectly. In congruence with the purpose of invalidating the knowledge or random guessing principle, such a strategy is taken that we find P^*_R which makes the resulting estimate of Index k* maximal. In so doing, we define entropy $\hat{H}^*$ such that

$$\hat{H}^* = \log \hat{k}^* = -\sum_{i=1}^{m} \hat{p}(i|\overline{A})\cdot\log \hat{p}(i|\overline{A}) \ . \tag{4.10}$$

Defining P^*_i such that

$$P^*_i \begin{cases} = P_i & i \neq R \\ = P^*_R & i = R \ , \end{cases} \tag{4.11}$$

we can write for the estimate, $\hat{p}(i|\overline{A})$, such that

$$\hat{p}(i|\overline{A}) = P^*_i\Big[\sum_{i=1}^{m} P^*_i\Big]^{-1} \ . \tag{4.12}$$

Then we can rewrite (4.10) in such a way that

$$\hat{H}^* = -\Big[\sum_{s=1}^{m} P^*_s\Big]^{-1}\Big[\sum_{i=1}^{m} P^*_i\cdot\log P^*_i - \Big(\sum_{i=1}^{m} P^*_i\Big)\cdot\log \Big\{\sum_{s=1}^{m} P^*_s\Big\}\Big] \ . \tag{4.13}$$

We have for the partial derivative of $\hat{H}^*$ with respect to P^*_R such that

$$\frac{\partial\hat{H}^*}{\partial P^*_R} = \Big[\sum_{s=1}^{m} P^*_s\Big]^{-2}\Big[\sum_{i=1}^{m} P^*_i\cdot\log P^*_i - \Big(\sum_{s=1}^{m} P^*_s\Big)\cdot\log P^*_R\Big] \ , \tag{4.14}$$

and, setting this derivative equal to zero, we obtain

$$P^*_R = \prod_{i\neq R} P_i^{\,P_i[\sum_{s\neq R} P_s]^{-1}} \ , \tag{4.15}$$

or

$$\log P^*_R = \Big[\sum_{s\neq R} P_s\Big]^{-1} \sum_{i\neq R} P_i\cdot\log P_i \ . \tag{4.16}$$

It is obvious that Index k* introduced in the preceding paragraphs can be used for our purpose of searching the direction that each multiple-choice item should take, i.e., Equivalent Distractor Model or Informative Distractor Model. If Index k* turns out to be far less than m , then we must reject the hypothesis of Equivalent Distractor Model for that test item. If it assumes a value close to m , then we shall say that Equivalent Distractor Model should still be under consideration. In both cases, Informative Distractor Model stays among the possibilities.

It is noted that the traditional chi-square test with (m-2) degrees of freedom for the goodness of fit for the frequencies of the (m-1) wrong answers with the uniform distribution as the theoretical distribution may serve our purpose just as well, without using Index k*. In our pilot study, we applied it for the original data of 7,439 examinees, and the result is summarized in Appendix IV. Table A-4-1 presents the frequency distribution of test items for each of the eleven subtests with respect to the probability resultant from this chi-square test, for each of Levels 11, 12 and 13. As we can see in these tables, only 23, 22 and 21 test items indicate the acceptance of the respective uniform distributions, or the acceptance of Equivalent Distractor Model, for Levels 11, 12 and 13, respectively, even if we take as low a level of significance as 0.0005 . Table A-4-2 presents the item identification, the probability and the percentage of the correct answer of each of these 66 test items. We can see in this table that, if we raise the level of significance to 0.01, then all the items whose probabilities are marked with ** will be excluded, to make the total number of items as small as 45, i.e., 17 for Level 11, 11 for Level 12, and 17 for Level 13. If, further, we raise the level of significance up to 0.05, then the total number of items will be reduced to 36, i.e., 15 for Level 11, 7 for Level 12, and 14 for Level 13, excluding all the items whose probabilities are marked with * . This number is only 2.5 percent of the total number of items, 1,448. There are only 19 items whose probabilities are greater than 0.2, i.e., R-78, W1-21, W1-25, W3-39, M1-31, M1-38, M1-45 and M1-68 for Level 11, V-79, R-78, W1-29, W2-44, W3-53 and M1-95 for Level 12, and V-93, R-140, R-155, W2-44, and M1-107 for Level 13. Since it is unlikely that this result obtained upon the original data is substantially different from the one obtainable from our revised data, we must take it as the suggestion for the rejection of Equivalent Distractor Model. Note, however, that the chi-square test applied for our data has severe criteria, since, in most cases, our sample sizes are very large. It is interesting to note that, except for W1-25 (57.9%) of Level 11, V-79 (46.6%) and M1-95 (29.4%) of Level 12, and V-93 (49.9%), R-140 (33.8%), R-155 (35.7%) and M1-107 (40.6%) of Level 13, all the other 12 test items, whose probabilites are greater than 0.2, have greater percentages of the correct answer than the medians of the separate subtests, as is clear if we compare Table A-4-2 with Table 3-1.

Since the examinees who answered correctly were excluded from those used for the chi-square test, we can see that these test items whose probabilities are greater than 0.2 are based upon relatively small sample sizes. As a whole, however, sample sizes are large, and the chi-square test is very sensitive to small deviations from the assigned uniform distributions.

In contrast to the chi-square test, we can see from (4.2) that the estimated Index k* is insensitive to the sample size, because the sampling fluctuation participates in the resulting estimate only through the computation of the proportions, P_i. Thus, whether it is right or wrong, if we wish to ignore the sampling fluctuations of the proportions, then the values of estimated Index k* can be comparable across different sample sizes.

Table 4-1 presents the frequency distribution of the items of each of the ten subtests, excluding Subtest L1, which consists of five-alternative test items, with respect to the resultant values of the estimated Index k*, for each of Levels 11, 12 and 13. The corresponding result for Subtest L1 is presented, separately, as Table 4-2, for all the three levels. We can see in Table 4-1 that the configurations of these frequencies are similar across the three levels, with the range of the estimated Index k*, 2.25 through 4.00, for each level. This is also the case with Subtest L1, with the range of estimated Index k* 2.25 through 4.50 for most items, as is shown in

TABLE 4-1

Frequency Distribution of Four-Alternative Items with Respect to Index k* for Each of the Ten Subtests. The Range of Index k* Is Greater Than or Equal to the Lower End and Less Than the Upper End of Each Interval.

Level 11

	Range of Index k*	Subtest										Total
		V	R	L2	L3	L4	W1	W2	W3	M1	M2	
1	1.00 - 1.25											0
2	1.25 - 1.50											0
3	1.50 - 1.75											0
4	1.75 - 2.00											0
5	2.00 - 2.25											0
6	2.25 - 2.50		1	2				1				4
7	2.50 - 2.75	1	2	7	1		1			1		13
8	2.75 - 3.00	3	2	6	8	3		2	1			25
9	3.00 - 3.25	6	6	10	12	8	1	2	5	3	1	54
10	3.25 - 3.50	3	13	8	7	12	4	1	9	4	1	62
11	3.50 - 3.75	11	13	6	8	6	7	4	12	7	12	86
12	3.75 - 4.00	19	37	1	4	3	23	16	29	27	15	174
	Total	43	74	40	40	32	36	26	56	42	29	418

Level 12

	Range of Index k*	Subtest										Total
		V	R	L2	L3	L4	W1	W2	W3	M1	M2	
1	1.00 - 1.25											0
2	1.25 - 1.50											0
3	1.50 - 1.75											0
4	1.74 - 2.00											0
5	2.00 - 2.25											0
6	2.25 - 2.50		1	4								5
7	2.50 - 2.75	2	1	8	1	1				1		14
8	2.75 - 3.00	2	4	6	8	3			5	2		30
9	3.00 - 3.25	4	10	7	8	8	2		8	6	1	54
10	3.25 - 3.50	6	9	8	10	11	4	3	11	5	3	70
11	3.50 - 3.75	10	18	5	11	5	6	8	16	8	10	97
12	3.75 - 4.00	22	33	4	4	4	28	17	19	23	17	171
	Total	46	76	42	42	32	40	28	59	45	31	441

Level 13

	Range of Index k*	Subtest										Total
		V	R	L2	L3	L4	W1	W2	W3	M1	M2	
1	1.00 - 1.25											0
2	1.25 - 1.50											0
3	1.50 - 1.75											0
4	1.75 - 2.00											0
5	2.00 - 2.25											0
6	2.25 - 2.50	2		3								5
7	2.50 - 2.75	3		7	2	1			1			14
8	2.75 - 3.00	2	5	7	4	2			2	2		24
9	3.00 - 3.25	1	5	10	11	7			7	4	1	46
10	3.25 - 3.50	11	7	8	10	10	5		10	9	5	75
11	3.50 - 3.75	10	24	5	11	6	7	10	21	12	7	113
12	3.75 - 4.00	19	37	3	5	6	29	18	18	21	19	175
	Total	48	78	43	43	32	41	28	59	48	32	452

TABLE 4-2

Frequency Distribution of Five-Alternative Items of Subtest L1, with Respect to Index k^*, for Levels 11, 12, and 13, Respectively.

	Range of Index k^*	Level			Total
		11	12	13	
1	1.00 - 1.25				0
2	1.25 - 1.50				0
3	1.50 - 1.75				0
4	1.75 - 2.00				0
5	2.00 - 2.25				0
6	2.25 - 2.50	4	4	1	9
7	2.50 - 2.75	5	1	2	8
8	2.75 - 3.00	4	4	4	12
9	3.00 - 3.25	2	2	8	12
10	3.25 - 3.50	5	11	6	22
11	3.50 - 3.75	9	7	5	21
12	3.75 - 4.00	4	5	11	20
13	4.00 - 4.25	5	6	4	15
14	4.25 - 4.50	4	5	4	13
15	4.50 - 4.75		1		1
16	4.75 - 5.00	1		3	4
	Total	43	46	48	137

Table 4-2. We notice in Table 4-1 that, for each of the three levels, the mode of the total frequency distribution is the highest category, 3.75 through 4.00 . If we examine the frequency distributions of separate subtests, however, we will notice that there are some variations among their configurations. Above all, it is noted that Subtests L2, L3 and L4 have different modes from the highest category, i.e., mostly either the category, 3.00 through 3.25, or the category, 3.25 through 3.50 . This tendency is also shared by Subtest L1, which has five-alternative multiple-choice test items, as is shown in Table 4-2.

Figure 4-1 presents the scatter diagram of all the four-alternative test items which are included in both Levels 11 and 12, with respect to the values of the estimated Index k*, and the corresponding scatter diagram for Levels 12 and 13. The numbers of the shared test items are 264 for Levels 11 and 12, and 323 for Levels 12 and 13. There is a tendency in both scatter diagrams that more items are located below the line of 45 degrees, which is drawn in the graphs, and this tendency is more conspicuous in the first one in which Levels 11 and 12 are compared. This indicates the existence of items whose values of the estimated Index k* are higher for the lower level and lower for the higher level. Similar scatter diagrams were drawn for each of the eleven subtests, including Subtest L1, and for each of the two pairs of levels, and the resultant twenty-two graphs are presented in Figures 4-2 and 4-3. We can see in

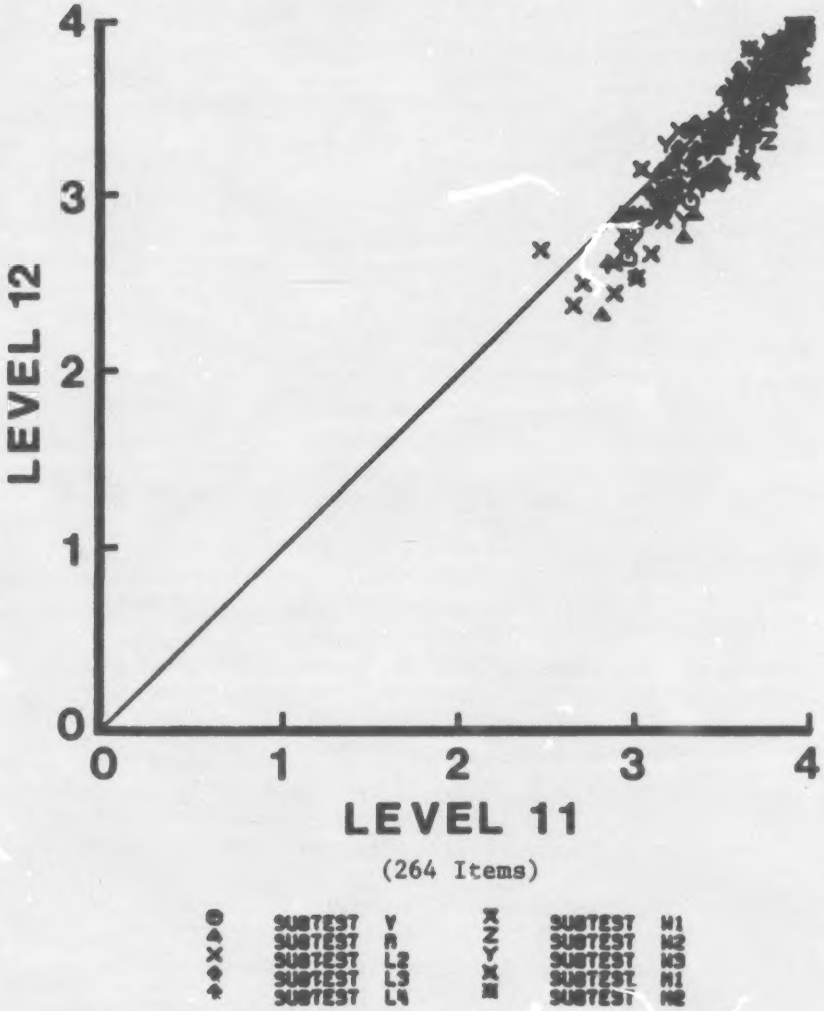


FIGURE 4-1

Comparison of the Values of Index k* for the Four-Alternative Items of the Eleven Subtests Administered to Two Adjacent Levels of Students: Levels 11 and 12.

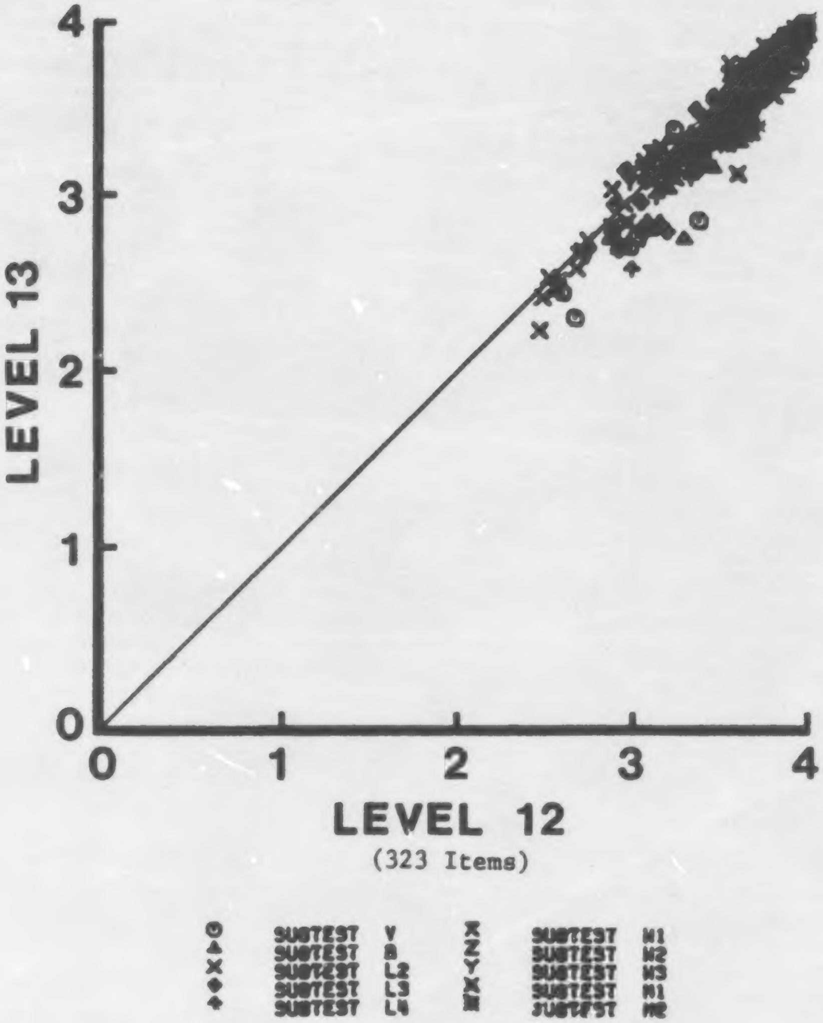
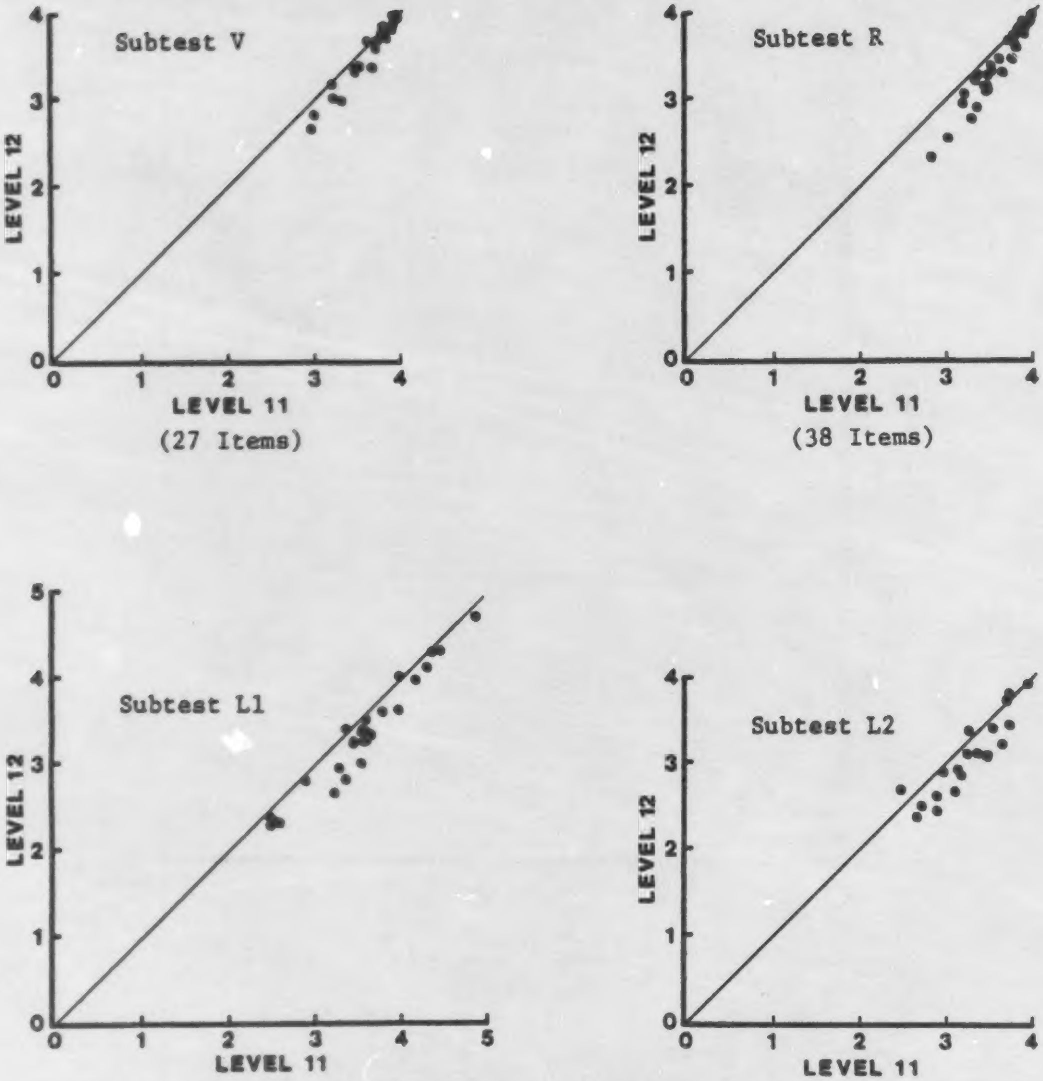


FIGURE 4-1 (Continued): Levels 12 and 13.

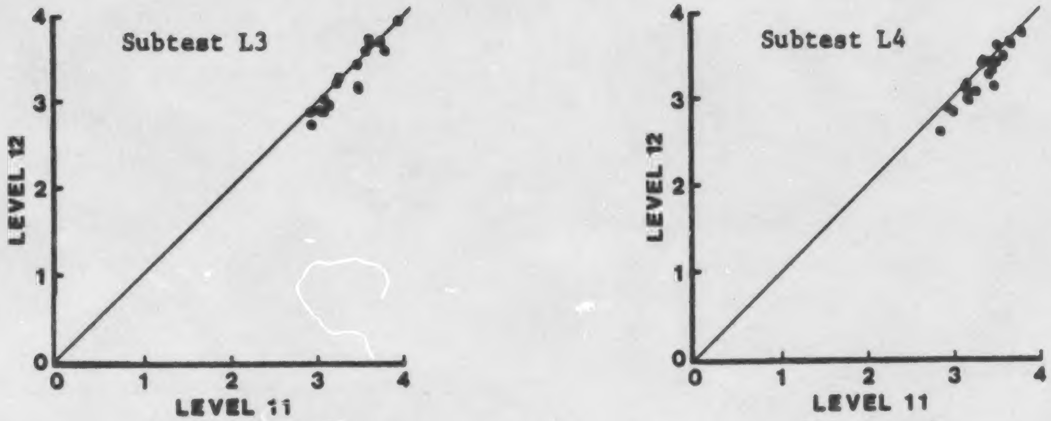


(27 Items) (38 Items)

(27 Items) (20 Items)

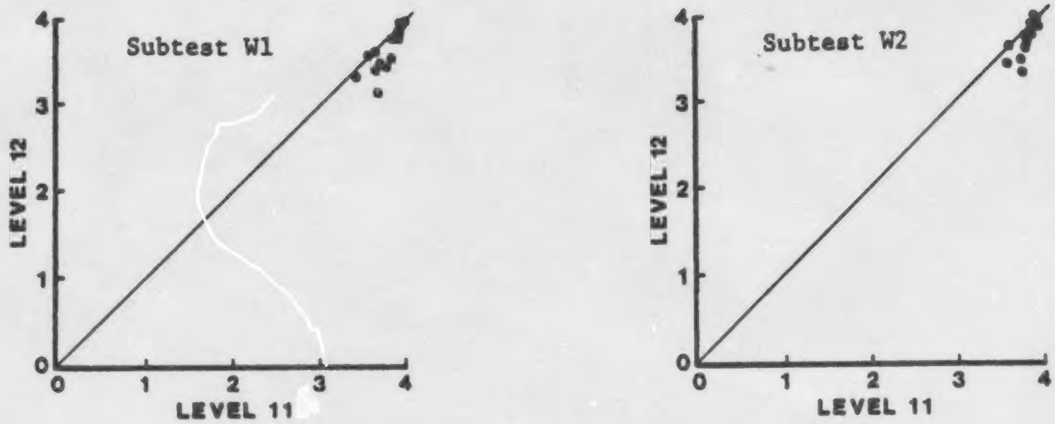
FIGURE 4-2

Comparison of the Values of Index k* for the Items of Each Subtest Administered to the Students of Both Levels 11 and 12.



(20 Items)

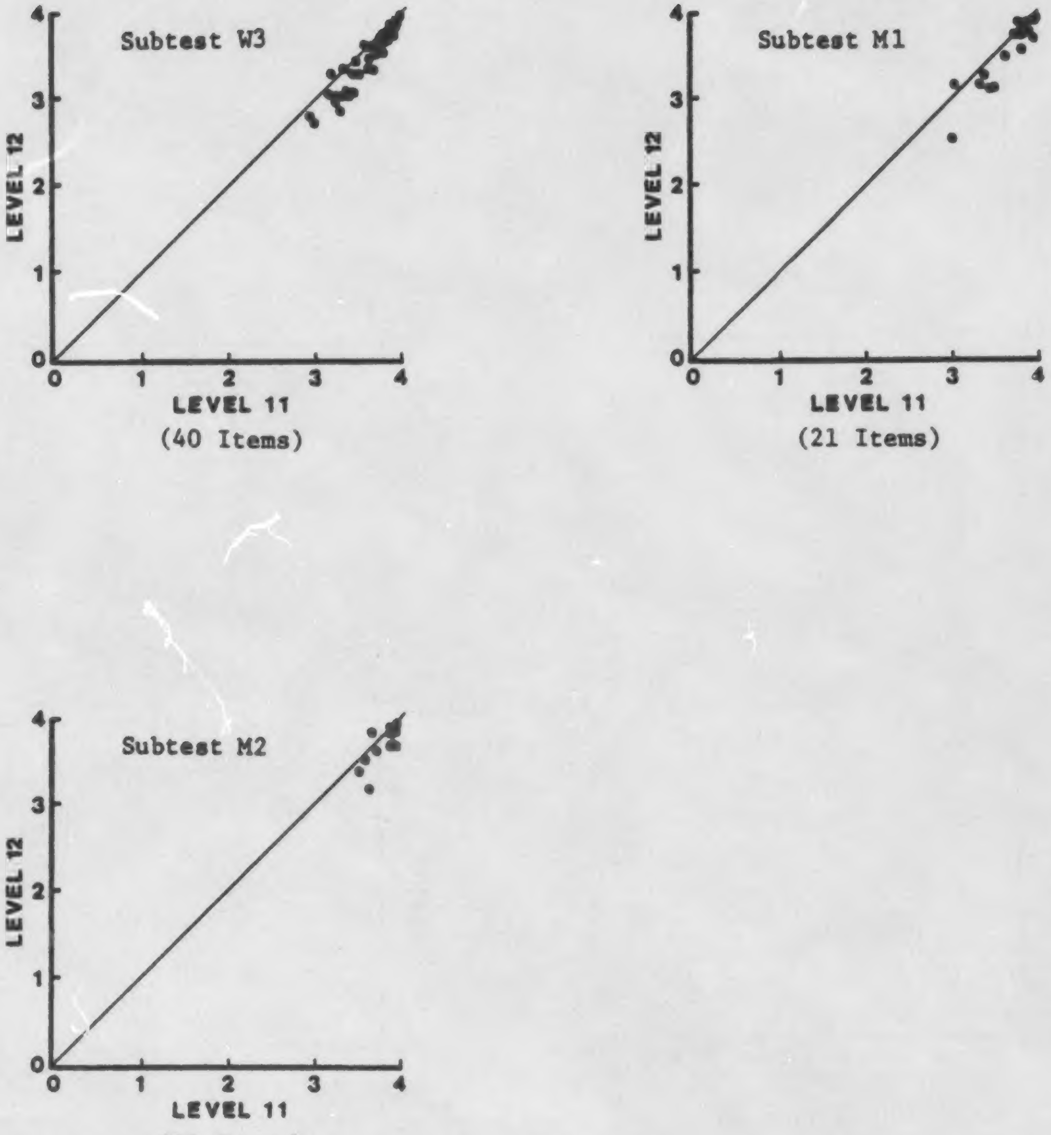
(22 Items)



(20 Items)

(14 Items)

FIGURE 4-2 (Continued): Levels 11 and 12.



(40 Items)

(21 Items)

(15 Items)

FIGURE 4-2 (Continued): Levels 11 and 12.

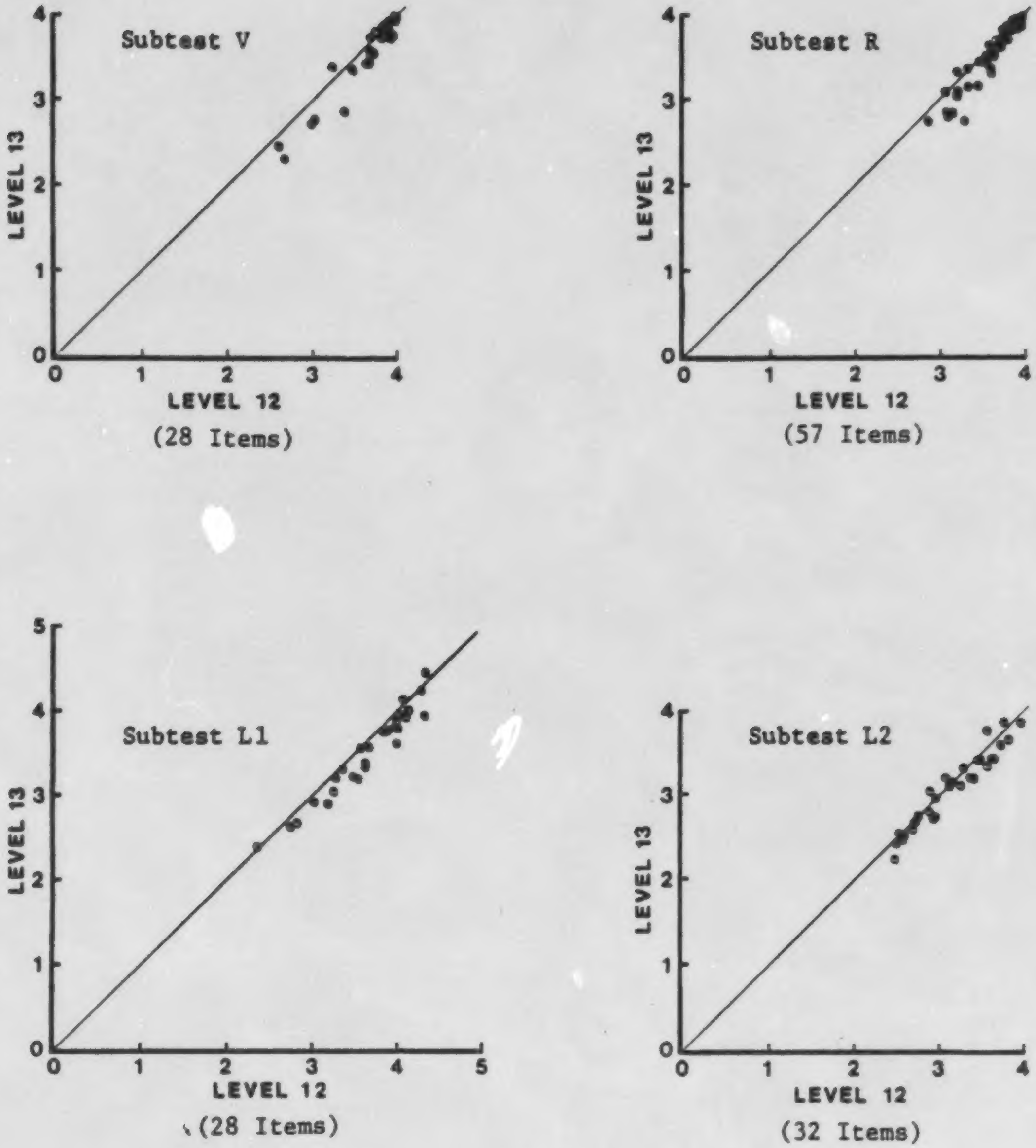
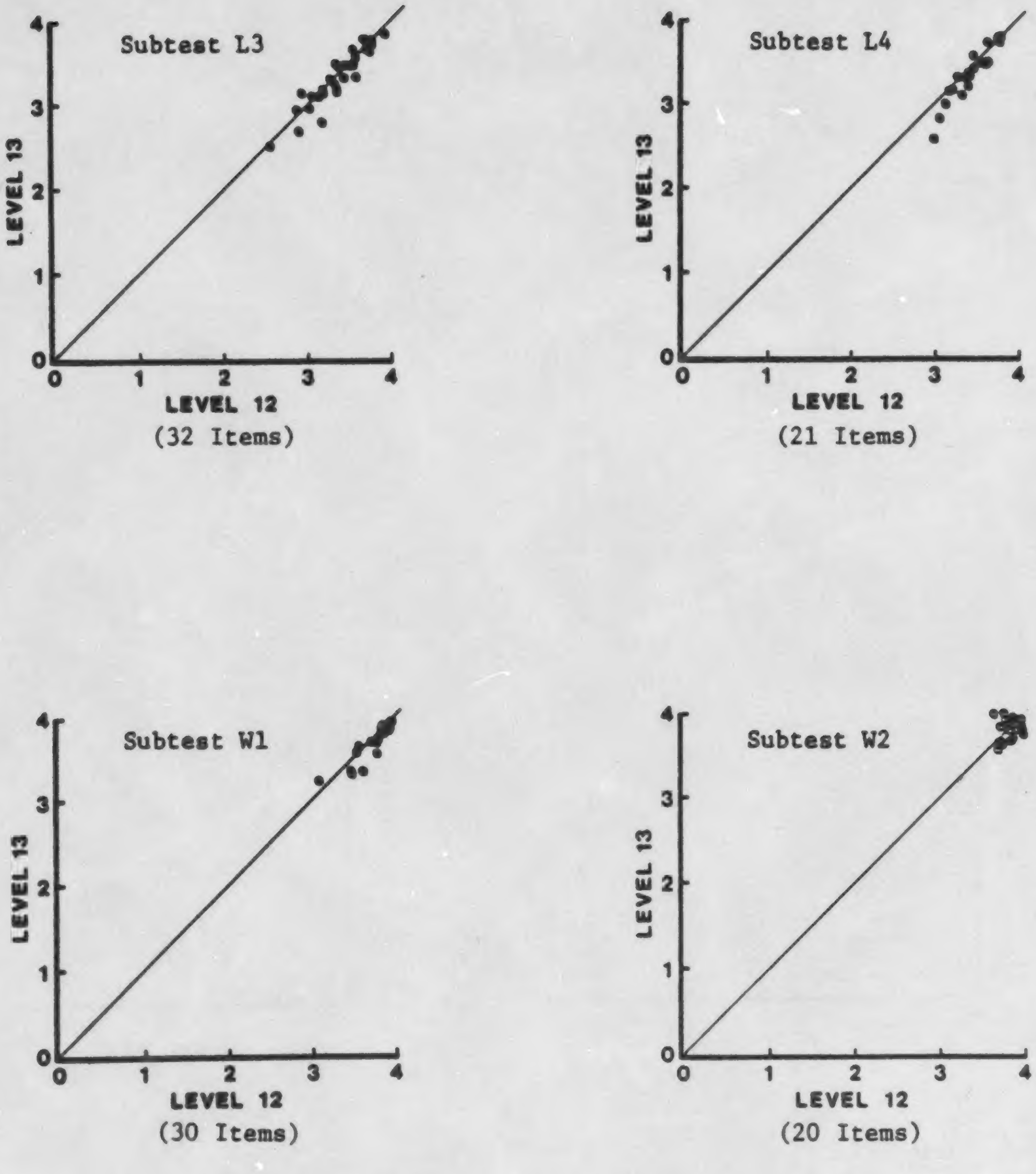


FIGURE 4-3

Comparison of the Values of Index k* for the Items of Each Subtest Administered to the Students of Both Levels 12 and 13.



(32 Items)

(21 Items)

(30 Items)

(20 Items)

FIGURE 4-3 (Continued): Levels 12 and 13.

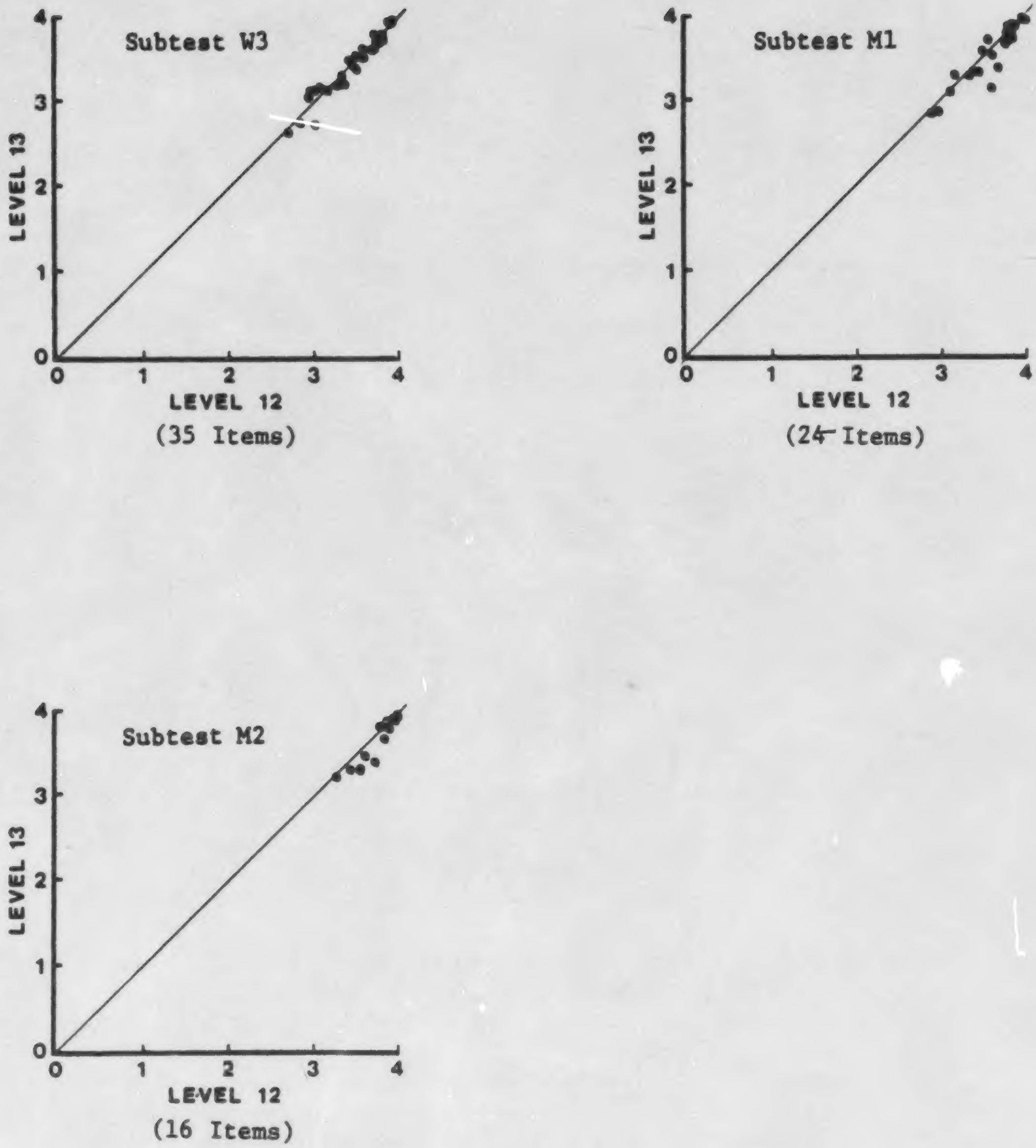


FIGURE 4-3 (Continued): Levels 12 and 13.

this figure that the above tendency exists in all the comparisons.

Table 4-3 presents the identification of each four-alternative test item whose estimated Index k* is 3.9 or greater, the value of the estimated Index k*, and the probability obtained as the result of the chi-square test, which was referred to earlier in this chapter. Note, however, that this probability was obtained upon the original data, while the estimated Index k* was obtained upon the revised data. The numbers of four-alternative test items whose estimated Index k* are greater than, or equal to, 3.9 are 78, 73 and 76 for Levels 11, 12 and 13, respectively. We can see there is an approximate correspondence between the probabilities which are greater than, or equal to, 0.1 and the estimated values of Index k* which are 3.9 or greater.

Eighty-two examples of the frequency distribution of the examinees with respect to their choices of an answer out of the four alternatives are presented as Figure 4-4. These test items are the first twenty-four items in each of the three levels in Table 4-3, added by those listed in the same table whose corresponding probabilities obtained as the result of the chi-square test are greater than, or equal to, 0.0005 . Thus we have 26, 27 and 29 four-alternative test items for Levels 11, 12 and 13, respectively, to make the total number of illustrated test items 82. In each histogram, also drawn by a dotted line is the estimated proportion, P_R^* , multiplied by the number of examinees who answered the item in one way or another, or the

TABLE 4-3

Probability Obtained by the Chi-Square Test (Original Data) against the Uniform Distribution for Items Whose Index k^*'s Are 3.9 or Greater (Revised Data). Items Are Arranged in the Descending Order of the Values of Index k^*.

Level 11

	Subtest & Item Number	Value of Index k^*	Probability
1	M1-068	3.99933	0.815
2	W1-021	3.99881	0.824
3	W3-039	3.99760	0.565
4	R -078	3.99728	0.889
5	W1-025	3.99669	0.514
6	M1-038	3.99635	0.669
7	M1-037	3.99544	0.121
8	M1-045	3.99404	0.246
9	W1-029	3.99401	0.147
10	V -061	3.99396	0.061
11	W3-053	3.99237	0.142
12	M1-031	3.99081	0.445
13	W3-040	3.99015	0.111
14	M1-063	3.98452	0.001
15	R -041	3.98349	0.000
16	R -047	3.98318	0.063
17	V -057	3.98067	0.000
18	R -088	3.97979	0.004
19	R -073	3.97964	0.002
20	M1-047	3.97779	0.000
21	W2-028	3.97392	0.000
22	M1-055	3.97082	0.000
23	W2-029	3.96987	0.012
24	W2-041	3.96974	0.001

TABLE 4-3 (Continued): Level 11.

	Subtest & Item Number	Value of Index k^*	Probability
25	W3-033	3.96878	0.000
26	R -090	3.96586	0.000
27	W1-043	3.96448	0.000
28	W1-045	3.96329	0.000
29	W3-063	3.96061	0.000
30	M2-034	3.95937	0.000
31	M2-037	3.95814	0.000
32	R -057	3.95716	0.000
33	M2-042	3.95493	0.000
34	L2-058	3.95473	0.000
35	W3-028	3.95427	0.000
36	L3-049	3.95320	0.000
37	W1-020	3.95258	0.036
38	W3-050	3.95042	0.000
39	W1-046	3.94947	0.000
40	M2-043	3.94942	0.000
41	W2-045	3.94929	0.000
42	V -058	3.94799	0.000
43	W3-060	3.94409	0.000
44	W2-043	3.94234	0.000
45	V -063	3.94127	0.000
46	V -048	3.93675	0.000
47	W1-047	3.93600	0.000
48	R -031	3.93572	0.000
49	R -042	3.93549	0.081
50	W3-070	3.93457	0.000
51	W1-019	3.93369	0.000
52	W1-028	3.93036	0.000
53	V -049	3.92777	0.000

TABLE 4-3 (Continued): Level 11.

	Subtest & Item Number	Value of Index k^*	Probability
54	V -052	3.92642	0.000
55	R -080	3.92629	0.000
56	W1-037	3.92554	0.000
57	W3-069	3.92481	0.000
58	W3-030	3.92469	0.000
59	V -054	3.92412	0.000
60	W3-078	3.92375	0.000
61	W1-040	3.92360	0.000
62	M1-069	3.92333	0.000
63	W3-031	3.91676	0.000
64	M2-048	3.91646	0.000
65	W2-044	3.91415	0.000
66	W1-031	3.91142	0.000
67	R -096	3.91013	0.000
68	R -097	3.91012	0.000
69	W1-038	3.90848	0.000
70	W1-035	3.90740	0.000
71	R -030	3.90710	0.000
72	R -032	3.90658	0.000
73	M1-061	3.90597	0.000
74	W3-044	3.90444	0.000
75	W1-018	3.90391	0.000
76	W3-047	3.90286	0.000
77	R -069	3.90207	0.000
78	M2-045	3.90191	0.000

TABLE 4-3 (Continued): Level 12.

	Subtest & Item Number	Value of Index k^*	Probability
1	M1-095	3.99933	0.660
2	R -078	3.99512	0.584
3	W1-029	3.99508	0.289
4	V -079	3.99508	0.209
5	W3-053	3.99274	0.383
6	V -057	3.99234	0.048
7	M1-074	3.99209	0.171
8	W2-044	3.99194	0.261
9	W1-053	3.99092	0.022
10	M2-059	3.98762	0.007
11	R -131	3.98608	0.018
12	V -049	3.98370	0.009
13	W2-048	3.98333	0.005
14	V -083	3.98260	0.000
15	W2-059	3.98129	0.004
16	V -052	3.97931	0.001
17	M2-060	3.97886	0.000
18	R -127	3.97850	0.001
19	R -088	3.97762	0.009
20	W3-083	3.97637	0.000
21	R -105	3.97371	0.001
22	W2-047	3.97358	0.000
23	W1-028	3.97239	0.001
24	W3-098	3.96896	0.000

TABLE 4-3 (Continued): Level 12.

	Subtest & Item Number	Value of Index k^*	Probability
25	R -116	3.96849	0.000
26	M2-054	3.96644	0.000
27	V -061	3.96614	0.000
28	M1-084	3.96585	0.000
29	R -128	3.96546	0.000
30	W2-045	3.96525	0.001
31	R -103	3.96483	0.000
32	W3-050	3.96465	0.001
33	W1-043	3.96262	0.000
34	M2-066	3.96052	0.000
35	W1-063	3.95866	0.000
36	W1-054	3.95746	0.000
37	L3-049	3.95658	0.000
38	V -076	3.95600	0.000
39	M1-068	3.95557	0.016
40	W1-045	3.95499	0.000
41	L2-058	3.95318	0.000
42	M1-088	3.95215	0.000
43	W1-067	3.95203	0.000
44	W1-062	3.95174	0.000
45	W1-057	3.94713	0.000
46	W3-093	3.94622	0.000
47	R -134	3.93945	0.000
48	W1-046	3.93684	0.000
49	M2-069	3.93447	0.000
50	R -133	3.92976	0.000
51	R -123	3.92899	0.000
52	R -073	3.92824	0.000
53	M2-046	3.92556	0.000

TABLE 4-3 (Continued): Level 12.

	Subtest & Item Number	Value of Index k^*	Probability
54	R -132	3.92516	0.000
55	W3-070	3.92434	0.000
56	V -058	3.92251	0.000
57	R -124	3.92229	0.000
58	W2-049	3.92155	0.000
59	W1-066	3.92136	0.000
60	M1-063	3.91931	0.000
61	R -122	3.91425	0.000
62	V -063	3.91414	0.000
63	W1-061	3.91392	0.000
64	W2-060	3.91359	0.000
65	R -092	3.91305	0.000
66	W3-063	3.90916	0.000
67	R -129	3.90905	0.000
68	M1-061	3.90706	0.000
69	R -090	3.90414	0.000
70	M1-057	3.90347	0.000
71	R -118	3.90282	0.000
72	W1-051	3.90055	0.000
73	W2-037	3.90046	0.000

TABLE 4-3 (Continued): Level 13.

	Subtest & Item Number	Value of Index k^*	Probability
1	V -093	3.99961	0.851
2	W2-044	3.99953	0.915
3	R -140	3.99820	0.403
4	M1-107	3.99691	0.280
5	R -155	3.99662	0.249
6	W3-083	3.99598	0.186
7	M2-075	3.99525	0.050
8	V -061	3.99401	0.122
9	W1-078	3.99350	0.066
10	W2-048	3.99311	0.096
11	R -142	3.99297	0.021
12	V -079	3.99110	0.063
13	R -131	3.99105	0.098
14	M1-105	3.98949	0.134
15	M2-060	3.98035	0.000
16	W2-066	3.97878	0.000
17	M2-082	3.97836	0.000
18	W2-067	3.97714	0.000
19	W1-075	3.97500	0.000
20	W1-053	3.97449	0.000
21	R -150	3.97339	0.000
22	R -103	3.97322	0.002
23	M2-078	3.97137	0.000
24	W3-070	3.97129	0.089

TABLE 4-3 (Continued): Level 13.

	Subtest & Item Number	Value of Index k^*	Probability
25	W3-098	3.96820	0.000
26	M1-118	3.96694	0.000
27	M1-088	3.96559	0.001
28	W1-069	3.96428	0.000
29	W3-123	3.96322	0.000
30	R -092	3.96306	0.032
31	M1-120	3.96178	0.000
32	R -088	3.95848	0.001
33	R -154	3.95847	0.000
34	W1-054	3.95749	0.000
35	W2-045	3.95585	0.001
36	W3-093	3.95506	0.000
37	M2-081	3.95472	0.000
38	R -122	3.95396	0.000
39	W1-077	3.95357	0.000
40	W3-105	3.95244	0.000
41	R -116	3.95196	0.000
42	M2-059	3.95147	0.000
43	M1-084	3.95103	0.000
44	R -128	3.95023	0.000
45	W3-114	3.94981	0.000
46	R -147	3.94961	0.000
47	M1-102	3.94521	0.000
48	W3-112	3.94497	0.000
49	W1-045	3.94399	0.000
50	M1-095	3.94364	0.000
51	W1-067	3.94181	0.000
52	V -083	3.94166	0.000
53	M1-074	3.94141	0.010

TABLE 4-3 (Continued): Level 13.

	Subtest & Item Number	Value of Index k^*	Probability
54	W2-063	3.94079	0.000
55	R -100	3.94043	0.000
56	V -092	3.94017	0.000
57	W2-064	3.93921	0.000
58	V -078	3.93713	0.000
59	W1-063	3.93531	0.000
60	M2-069	3.93418	0.000
61	M1-112	3.93004	0.000
62	W2-047	3.92916	0.000
63	W1-060	3.92723	0.000
64	W2-041	3.92541	0.000
65	W2-059	3.92402	0.000
66	W1-068	3.91678	0.000
67	R -129	3.91674	0.000
68	R -118	3.91623	0.000
69	R -133	3.91549	0.000
70	W1-066	3.91517	0.000
71	W1-064	3.90854	0.000
72	W1-046	3.90724	0.000
73	W2-068	3.90657	0.000
74	M2-066	3.90498	0.000
75	V -066	3.90441	0.000
76	R -130	3.90411	0.000

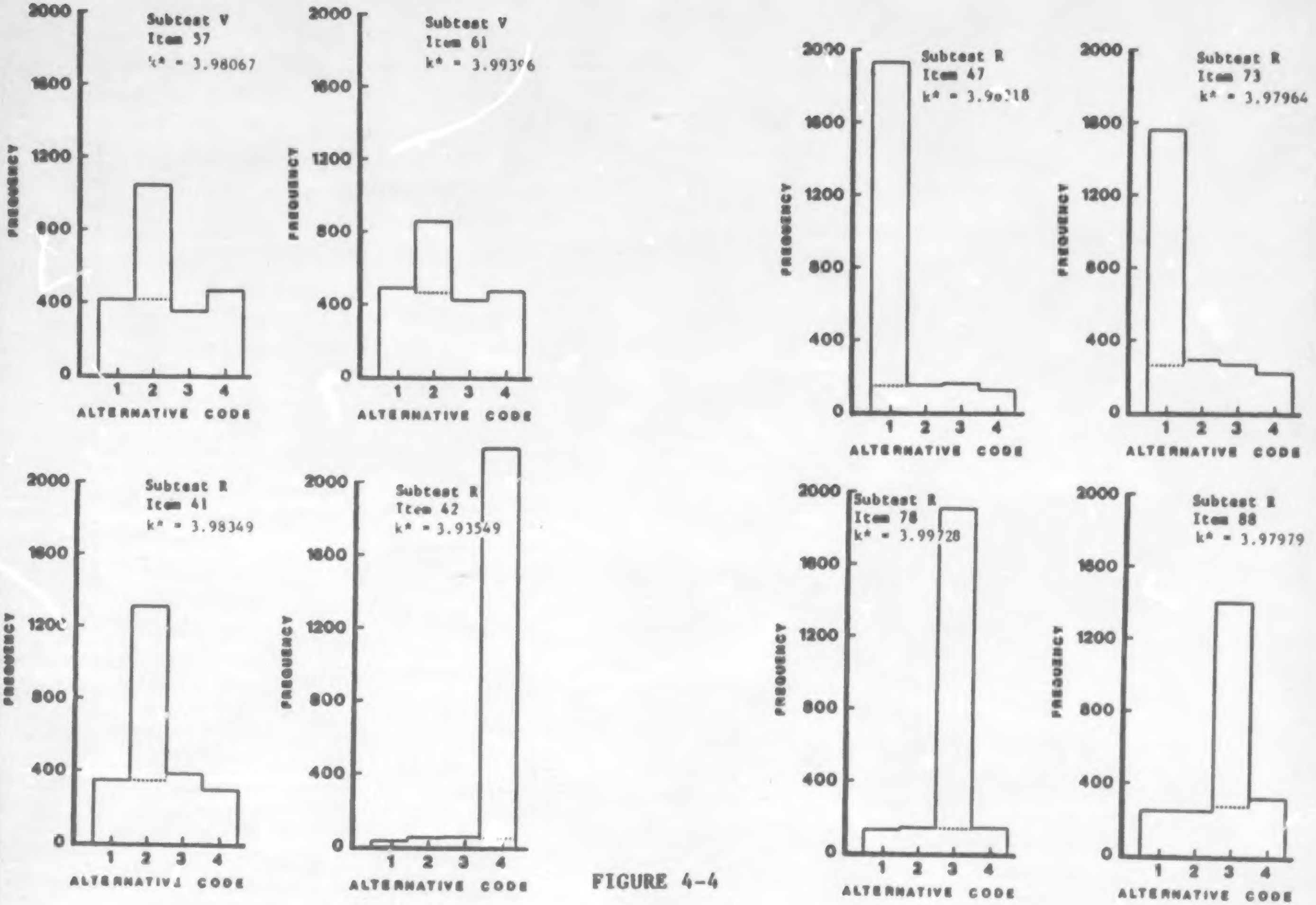


FIGURE 4-4

Frequency Distribution of the Examinees of Each of the Three Levels with Respect to Their Responses to Each Test Item Whose Index k* Is 3.9 or Greater, with the Estimated Proportion of the Examinees Guessing Correctly (Dotted Line).

Level 11

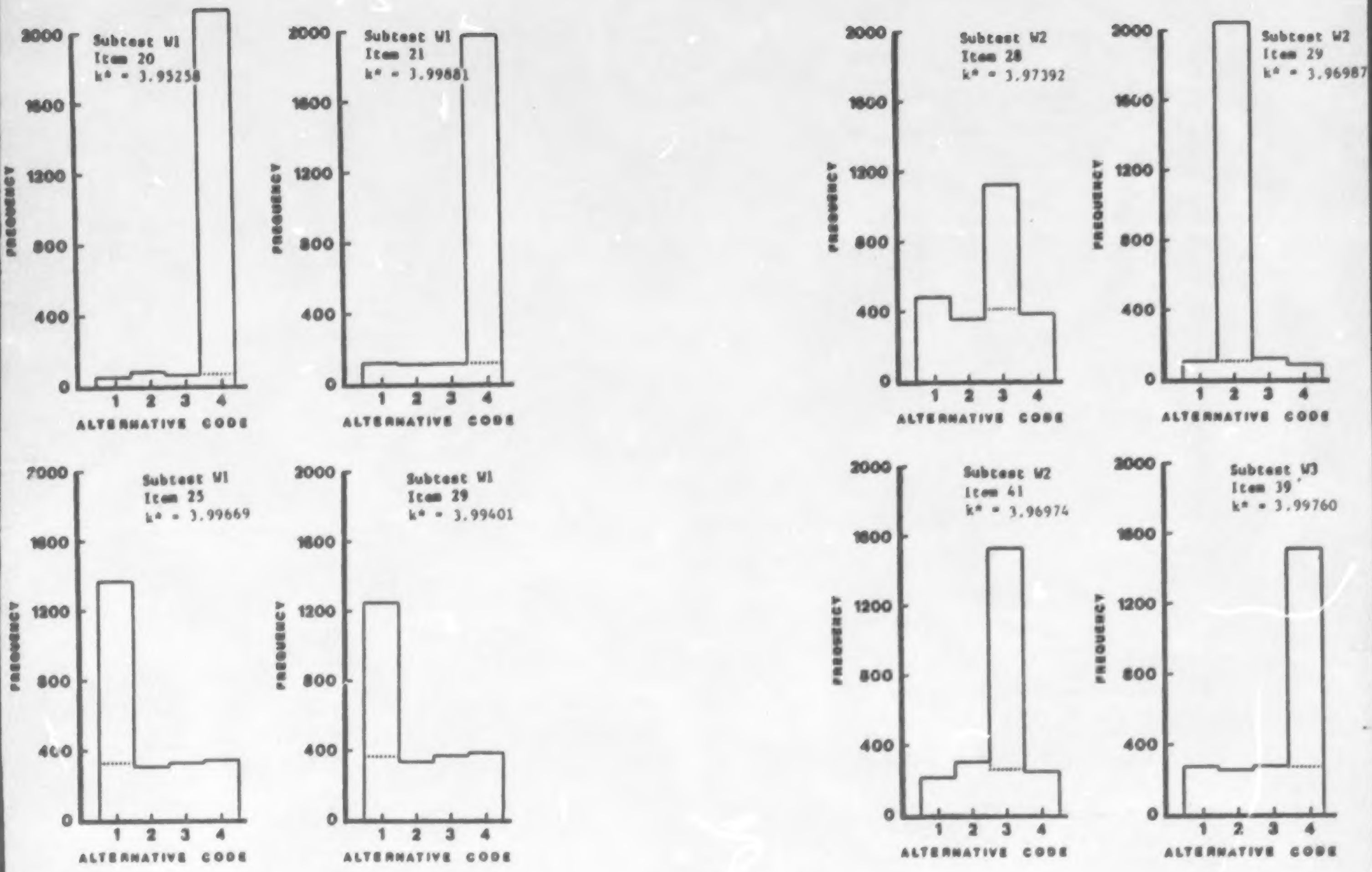


FIGURE 4-4 (Continued): Level 11.

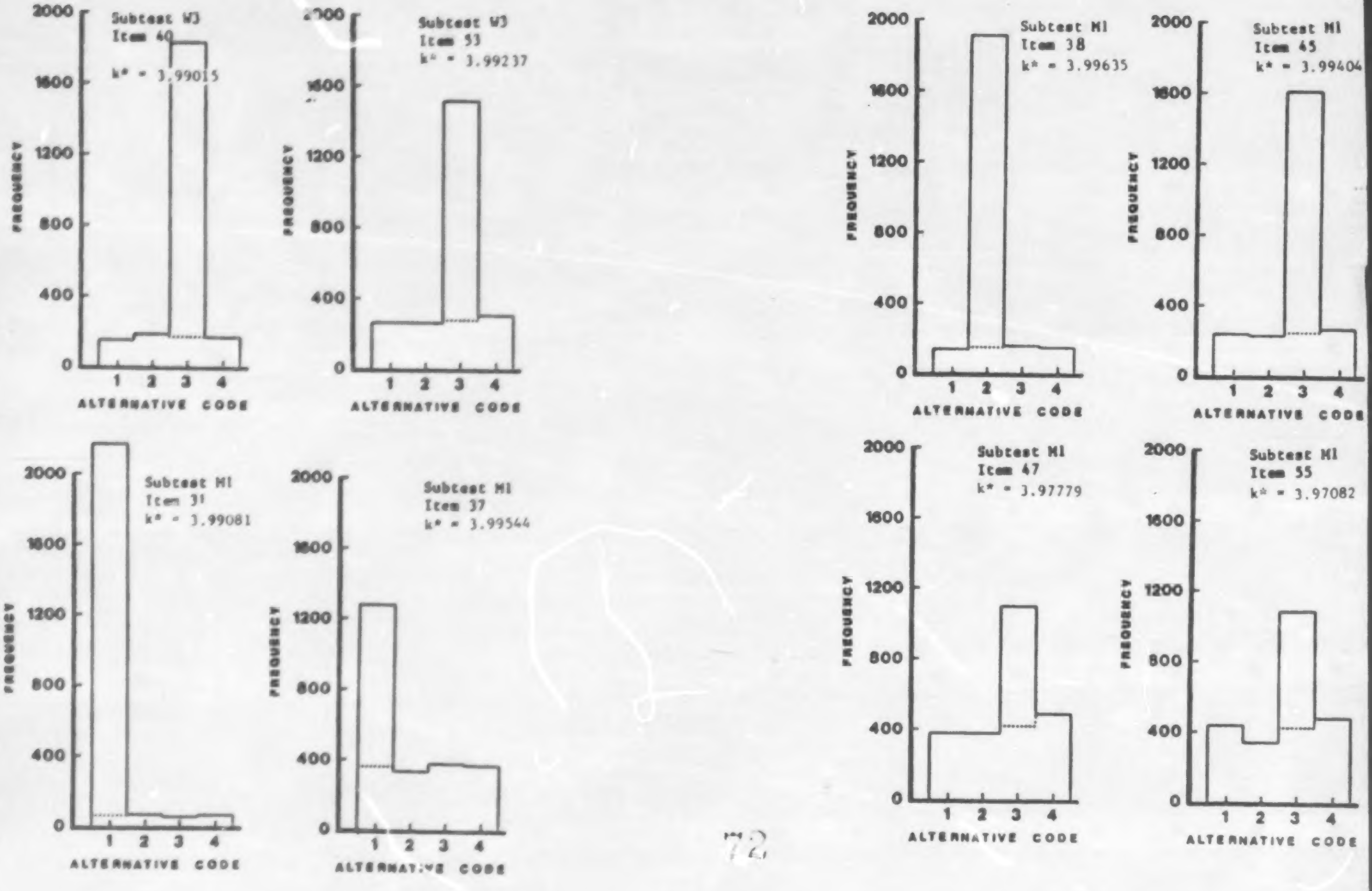


FIGURE 4-4 (Continued): Level 11.

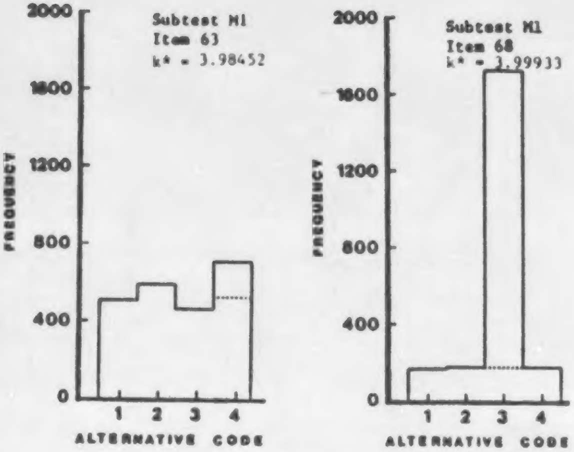
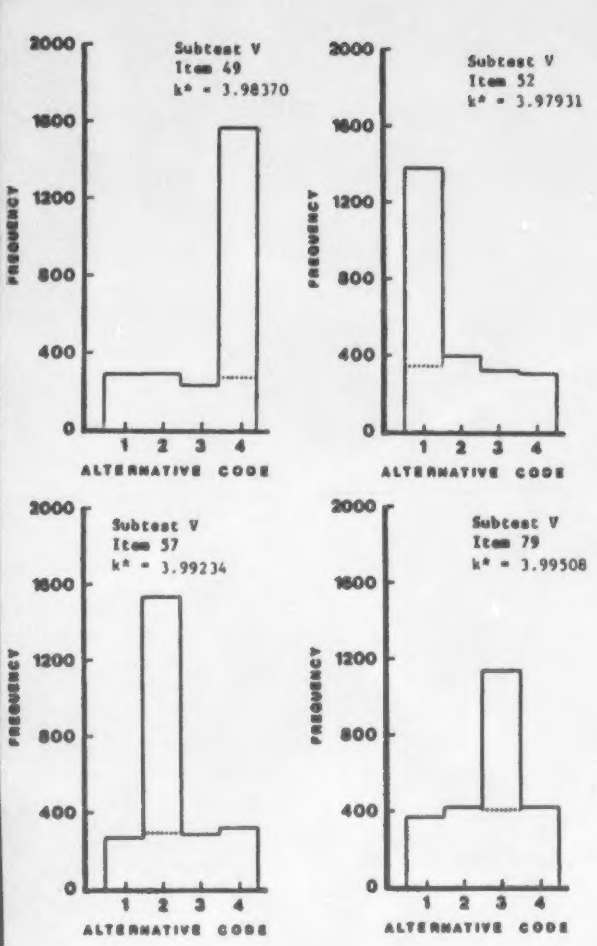


FIGURE 4-4 (Continued): Level 11.



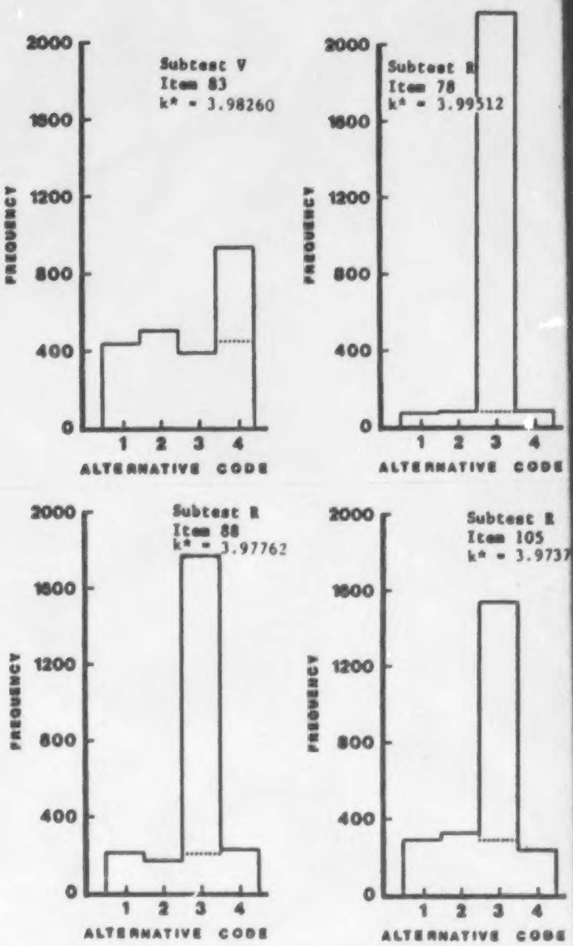


FIGURE 4-4 (Continued): Level 12.

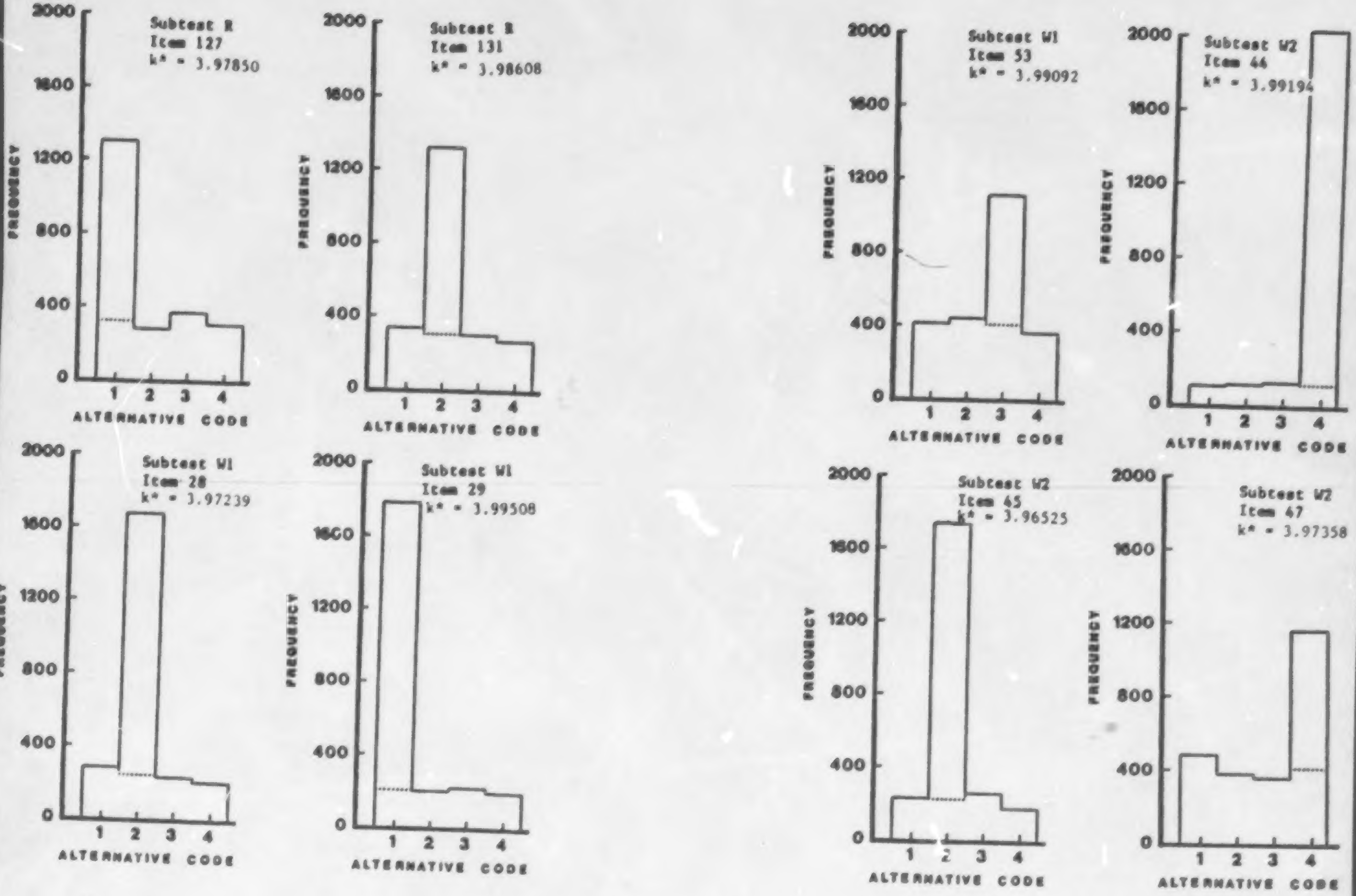


FIGURE 4-4 (Continued): Level 12.

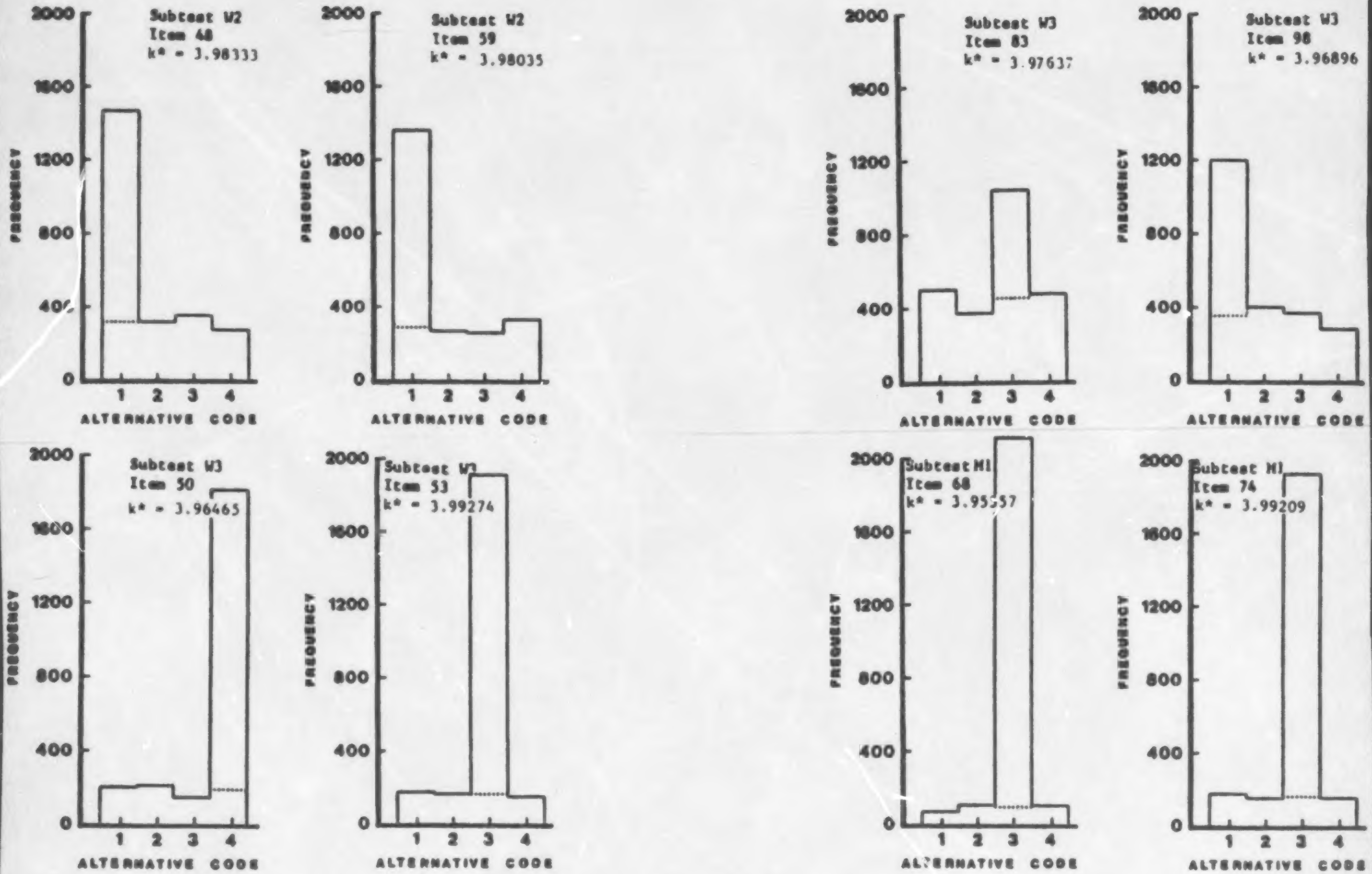


FIGURE 4-4 (Continued): Level 12.

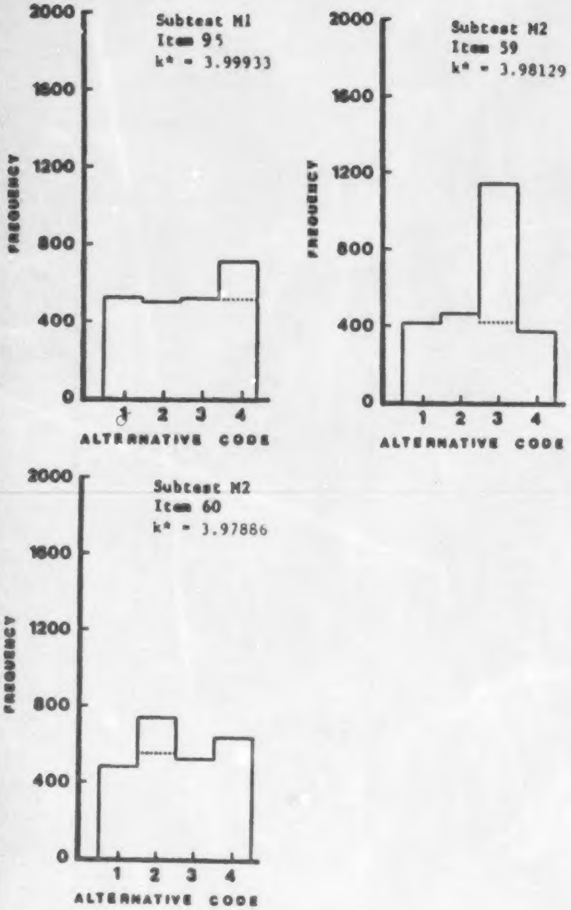


FIGURE 4-4 (Continued): Level 12.

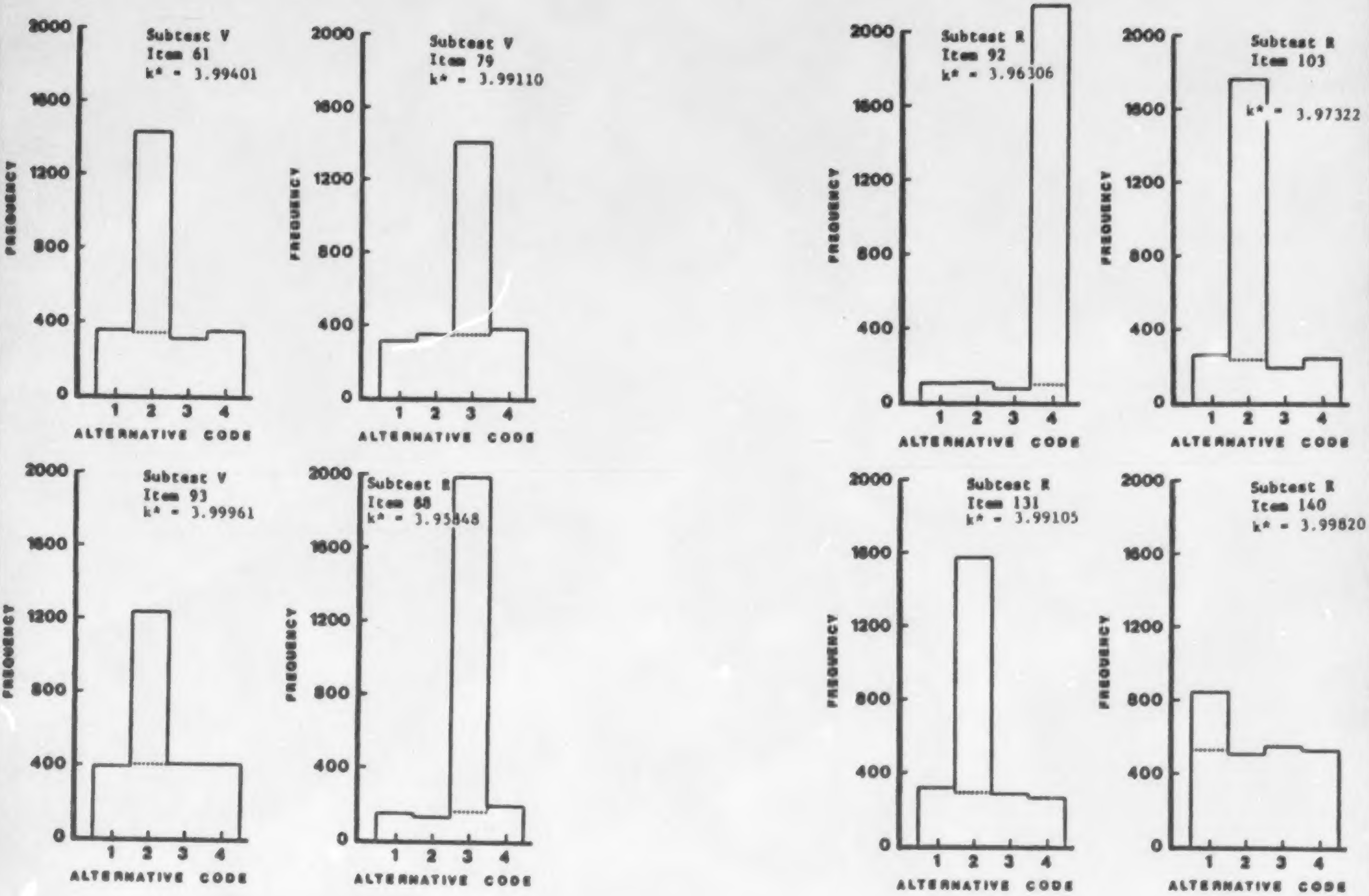


FIGURE 4-4 (Continued): Level 13.

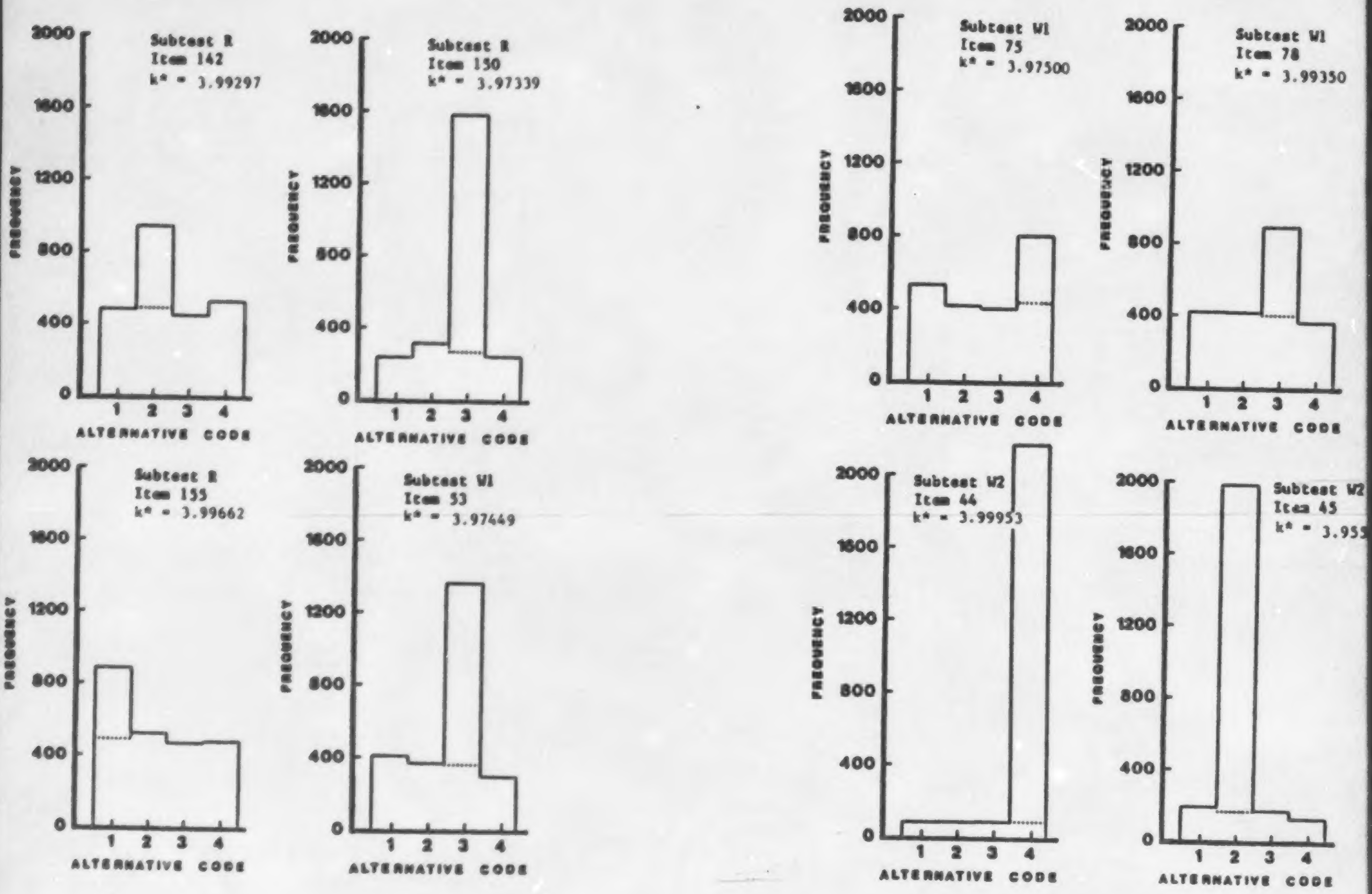


FIGURE 4-4 (Continued): Level 13.

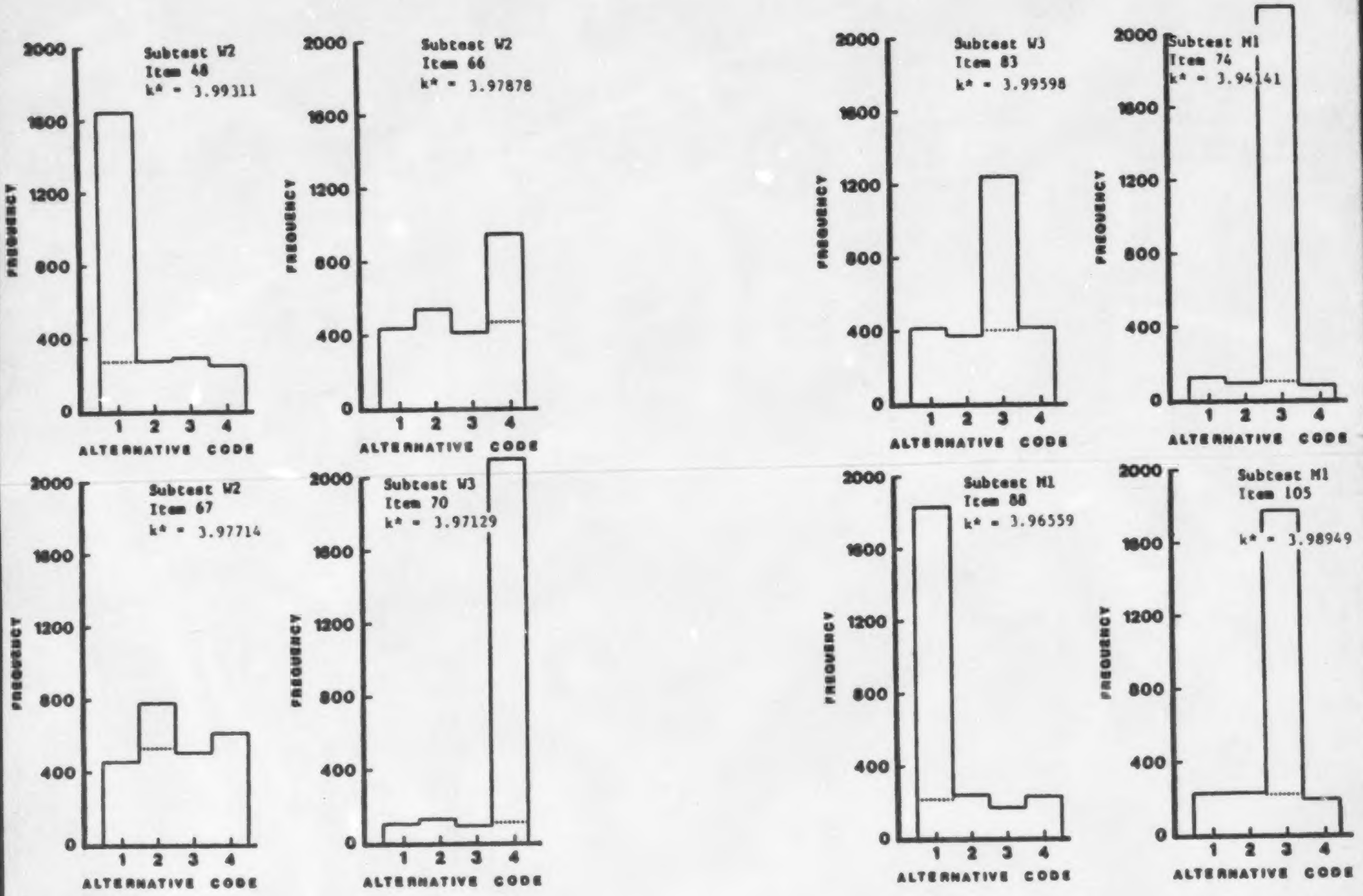


FIGURE 4-4 (Continued): Level 13.

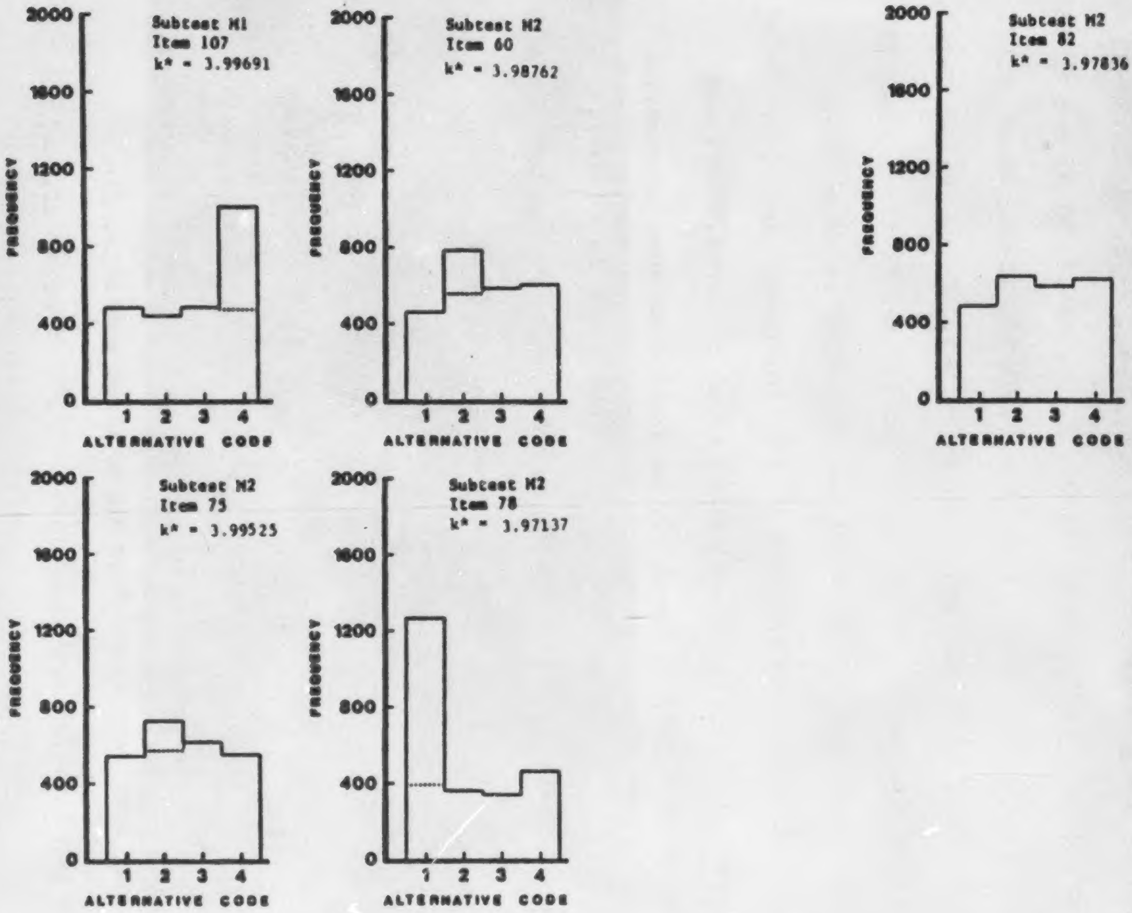


FIGURE 4-4 (Continued): Level 13.

total number of examinees subtracted by the number of those who did not answer the item at all. We can see in this figure that most of these histograms are close to rectangles, if we replace the frequency for the correct answer by the height indicated by the dotted line in each histogram. It is especially true if we focus our attention to those items whose estimates of Index k* are greater than 3.99, i.e., V-61, R-78, W1-21, W1-25, W1-29, W3-39, W3-40, W3-53, M1-31, M1-37, M1-38, M1-45 and M1-68 of Level 11, V-57, V-79, R-78, W1-29, W1-53, W2-44, W3-53, M1-74 and M1-95 of Level 12, and V-61, V-79, V-93, R-131, R-140, R-142, R-155, W1-78, W2-44, W2-48, W3-83, M1-107 and M2-75 of Level 13. On the other hand, if we shift our attention to those whose values of the estimated Index k* are relatively low, like W2-28 of Level 11 (k*=3.97392), W3-83 of Level 12 (k*=3.97637) and M2-78 (k*=3.97137) of Level 13, we shall find some substantial deviations from rectangularity. For the group as a whole, however, these histograms are close enough to rectangles, as we expect from the high values of estimated Index k*.

It should be noted that among these eighty-two four-alternative test items we find no items from Subtest L2, L3 or L4. These subtests are all developed for measuring language skills. In fact, in the total group of 227 test items, whose values of the estimated Index k* are greater than 3.9, there are only four test items from these three subtests, i.e., L2-58 (k*=3.95473) and L3-49 (k*=3.95320) of Level 11, and

L3-49 (k*=3.95658) and L2-58 (k*=3.95318) of Level 12, which are actually two items shared by both Levels 11 and 12. This result is no surprise, considering the fact that for these subtests the modes of the estimated Index k* are much lower than those of the other subtests, as we have seen earlier in Table 4-1. A close examination of the contents of the test items of these four subtests, including Subtest L1, and their results of analysis reveals the following facts.

(1) All the questions in these four language skill subtests are in the form of having the examinee find mistakes in spelling, capitalization, punctuation and usage, respectively.

(2) Unlike the test items in the other seven subtests, these items have "No mistakes" as the last alternative, and for most items this alternative has a high frequency, even when it is a wrong answer.

From these facts and the above results, it is obvious that Equivalent Distractor Model is not suitable for the items of the four subtests of language skills, including Subtest L1, which consists of five-alternative test items. For these items, Informative Distractor Model may be more appropriate.

It should also be noted that, except for a few items like V-61 and M1-63 of Level 11, M1-95 and M2-60 of Level 12, and R-140, R-155, W1-75, W1-78, W2-67, M2-60, M2-75 and M2-82 of Level 13, these items presented in Figure 4-4 are relatively easy items for the respective groups of examinees. In such a case, we cannot expect this type of frequency distribution to reveal the information each alternative has, even if the item belongs to Informative Distractor Model. If we exclude all the four-alternative test items which belong to one of the three subtests of language skills, there are 76 items out of 306 for Level 11, 71 items out of 325 for Level 12, and 76 out of 334 for Level 13, whose estimates of Index k* are 3.9 or greater. These numbers are interpreted as 24.8, 21.8 and 22.8 in percentage for Levels 11, 12 and 13, respectively, for the remaining seven subtests. They are by no means large numbers, however.

Table 4-4 presents the item identification, the value of estimated Index k*, and the probability obtained by the chi-square test for each of the four-alternative test items whose estimates of Index k* are the lowest, i.e., 3.0 or less. These items number 134, of which 42 are of Level 11, 49 are of Level 12, and the remaining 43 are of Level 13. We can see in this table that all the probabilities, which were obtained upon the original data, are less than 0.0005 . It is noted that there are many items from Subtest L2, L3, or L4, unlike those for the highest Index k* group, which are shown in Table 4-3. In fact, there are as many as 27 such items for Level 11, 31 for Level 12 and 26 for Level 13. Out of these test items, 7, 11 and 9 items

TABLE 4-4

Probability Obtained by the Chi-Square Test against the Uniform Distribution for Item Alternatives Whose Index k* Values Are 3.0 or Less. Items Are Arranged in the Descending Order of the Values of Index k* .

Level 11

	Subtest & Item Number	Value of Index k^*	Probability
1	V -060	2.98574	0.000
2	L3-038	2.98446	0.000
3	L2-053	2.98120	0.000
4	W2-022	2.98077	0.000
5	W2-026	2.96675	0.000
6	L4-038	2.96582	0.000
7	L3-053	2.96109	0.000
8	L3-041	2.95480	0.000
9	W3-043	2.95337	0.000
10	L3-033	2.94519	0.000
11	L3-022	2.94137	0.000
12	L3-052	2.93717	0.000
13	V -027	2.92684	0.000
14	L4-024	2.91176	0.000
15	L2-046	2.90829	0.000
16	L2-048	2.90430	0.000
17	L2-029	2.90014	0.000
18	L2-027	2.88903	0.000
19	L4-041	2.87004	0.000
20	R -063	2.83820	0.000
21	L2-023	2.83691	0.000
22	R -037	2.83280	0.000
23	L3-036	2.83258	0.000
24	V -034	2.78061	0.000

TABLE 4-4 (Continued): Level 11.

	Subtest & Item Number	Value of Index k^*	Probability
25	L3-020	2.77848	0.000
26	L2-054	2.72988	0.000
27	M1-032	2.72056	0.000
28	L2-035	2.71827	0.000
29	L2-034	2.70963	0.000
30	L2-040	2.67619	0.000
31	L3-025	2.65321	0.000
32	L2-025	2.63903	0.000
33	L2-028	2.63035	0.000
34	V -024	2.62287	0.000
35	R -043	2.61763	0.000
36	R -050	2.58812	0.000
37	L2-019	2.54288	0.000
38	W1-027	2.52326	0.000
39	L2-055	2.49083	0.000
40	L2-024	2.47698	0.000
41	R -034	2.45910	0.000
42	W2-023	2.31156	0.000

TABLE 4-4 (Continued): Level 12.

	Subtest & Item Number	Value of Index k^*	Probability
1	W3-072	2.99099	0.000
2	V -071	2.98965	0.000
3	L4-034	2.98524	0.000
4	L3-048	2.98363	0.000
5	M1-085	2.97425	0.000
6	R -067	2.96648	0.000
7	L3-069	2.96261	0.000
8	L2-068	2.95951	0.000
9	L2-056	2.95558	0.000
10	L2-053	2.92803	0.000
11	L3-046	2.92549	0.000
12	L3-053	2.92517	0.000
13	R -071	2.91927	0.000
14	L3-040	2.90879	0.000
15	L3-042	2.89935	0.000
16	L3-052	2.89917	0.000
17	L4-038	2.89777	0.000
18	W3-084	2.89085	0.000
19	W3-045	2.89001	0.000
20	L2-049	2.88979	0.000
21	M1-091	2.88760	0.000
22	L2-061	2.88622	0.000
23	R -106	2.87424	0.000
24	L4-043	2.85821	0.000

TABLE 4-4 (Continued): Level 12.

	Subtest & Item Number	Value of Index k^*	Probability
25	V -053	2.84096	0.000
26	W3-043	2.83073	0.000
27	R -065	2.78972	0.000
28	L3-041	2.75618	0.000
29	L2-077	2.75246	0.000
30	W3-063	2.75097	0.000
31	L2-055	2.72287	0.000
32	L2-080	2.70300	0.000
33	L2-041	2.70073	0.000
34	L2-078	2.68977	0.000
35	V -060	2.68016	0.000
36	L2-048	2.64858	0.000
37	L4-041	2.62909	0.000
38	V -084	2.61062	0.000
39	L2-070	2.58695	0.000
40	L3-072	2.58167	0.000
41	L2-063	2.57592	0.000
42	R -066	2.56920	0.000
43	M1-054	2.56018	0.000
44	L2-054	2.53272	0.000
45	L2-079	2.49724	0.000
46	L2-067	2.47992	0.000
47	L2-046	2.47486	0.000
48	L2-040	2.40572	0.000
49	R -063	2.34781	0.000

TABLE 4-4 (Continued): Level 13.

	Subtest & Item Number	Value of Index k^*	Probability
1	L3-054	2.98990	0.000
2	L2-068	2.97913	0.000
3	L3-052	2.97289	0.000
4	L4-071	2.92099	0.000
5	V -059	2.88186	0.000
6	R -109	2.87719	0.000
7	M1-085	2.87223	0.000
8	L2-085	2.87204	0.000
9	R -082	2.87129	0.000
10	M1-091	2.84972	0.000
11	L4-050	2.84871	0.000
12	L2-081	2.83841	0.000
13	L3-086	2.82643	0.000
14	R -083	2.82386	0.000
15	L3-055	2.82275	0.000
16	L2-061	2.81845	0.000
17	W3-084	2.79924	0.000
18	V -064	2.79228	0.000
19	W3-067	2.77642	0.000
20	R -087	2.77484	0.000
21	R -106	2.77466	0.000
22	L2-077	2.77141	0.000
23	L2-056	2.76178	0.000
24	L2-088	2.75648	0.000

TABLE 4-4 (Continued): Level 13.

	Subtest & Item Number	Value of Index k^*	Probability
25	L2-053	2.73676	0.000
26	V -071	2.73664	0.000
27	L3-053	2.71696	0.000
28	L2-055	2.71265	0.000
29	V -096	2.69742	0.000
30	L2-080	2.69277	0.000
31	W3-068	2.68418	0.000
32	L2-090	2.63891	0.000
33	L2-078	2.61208	0.000
34	L4-044	2.60493	0.000
35	L2-054	2.56223	0.000
36	L2-070	2.54443	0.000
37	L3-072	2.54169	0.000
38	V -099	2.50520	0.000
39	L2-063	2.49213	0.000
40	V -084	2.47745	0.000
41	L2-079	2.44666	0.000
42	V -060	2.32986	0.000
43	L2-067	2.25958	0.000

have the values of the estimated Index k* less than, or equal to, 2.6, respectively, for Levels 11, 12 and 13. Histograms similar in nature as those in Figure 4-4 are drawn for the frequency distributions of these 27 four-alternative test items, and are presented as Figure 4-5. We can see that these histograms, whose frequencies for the correct answer are replaced by the dotted lines, are by no means close to rectangles. There is no reason to accept Equivalent Distractor Models for these test items. Among them, there are such items as R-34 and R-50 for Level 11, L2-79 and L3-72 for Level 12, and L2-79 and L3-72 for Level 13 whose values of P^*_R multiplied by the respective numbers of examinees who answered in one way or another exceed the corresponding frequencies for the correct answer; in these cases, the use of Index k* itself is meaningless.

Figure 4-6 presents the frequency distributions of six four-alternative test items for each of the three levels, which were arbitrarily selected from those whose values of the estimated Index k* are greater than, or equal to, 3.0 and less than 3.6. They are samples from the four-alternative test items with intermediate values of the estimated Index k*. We can see that these histograms, in which the frequencies for the correct answer are replaced by the dotted lines, are also far from rectangles. They also include one item, i.e., L4-74 for Level 13, for which the dotted line exceeds the frequency

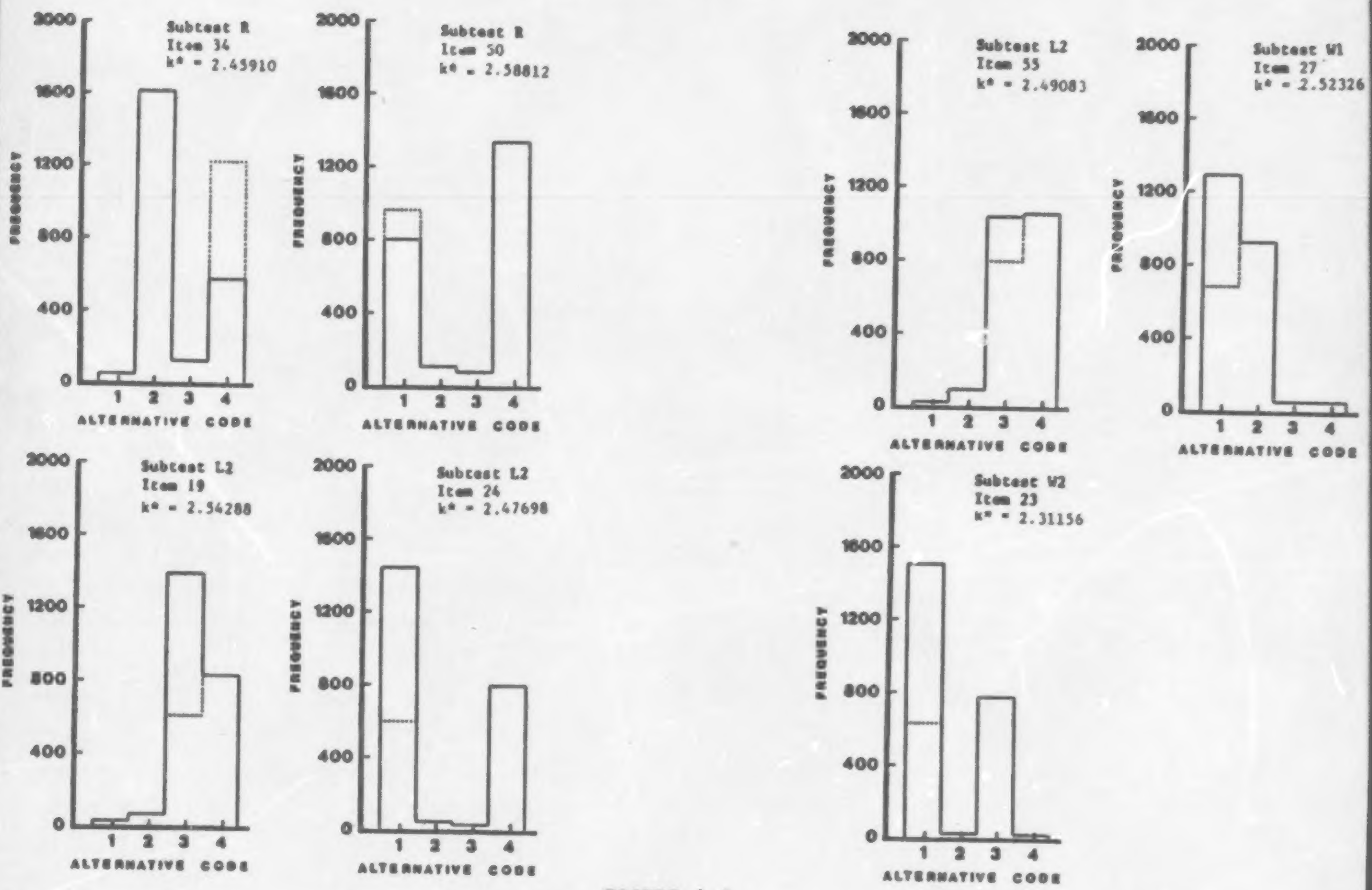


FIGURE 4-5

Frequency Distribution of the Examinees of Each of the Three Levels with Respect to Their Responses to Each Test Item Whose Index k* Is 2.6 or Less, with the Estimated Proportion of the Examinees Guessing Correctly (Dotted Line).

Level 11

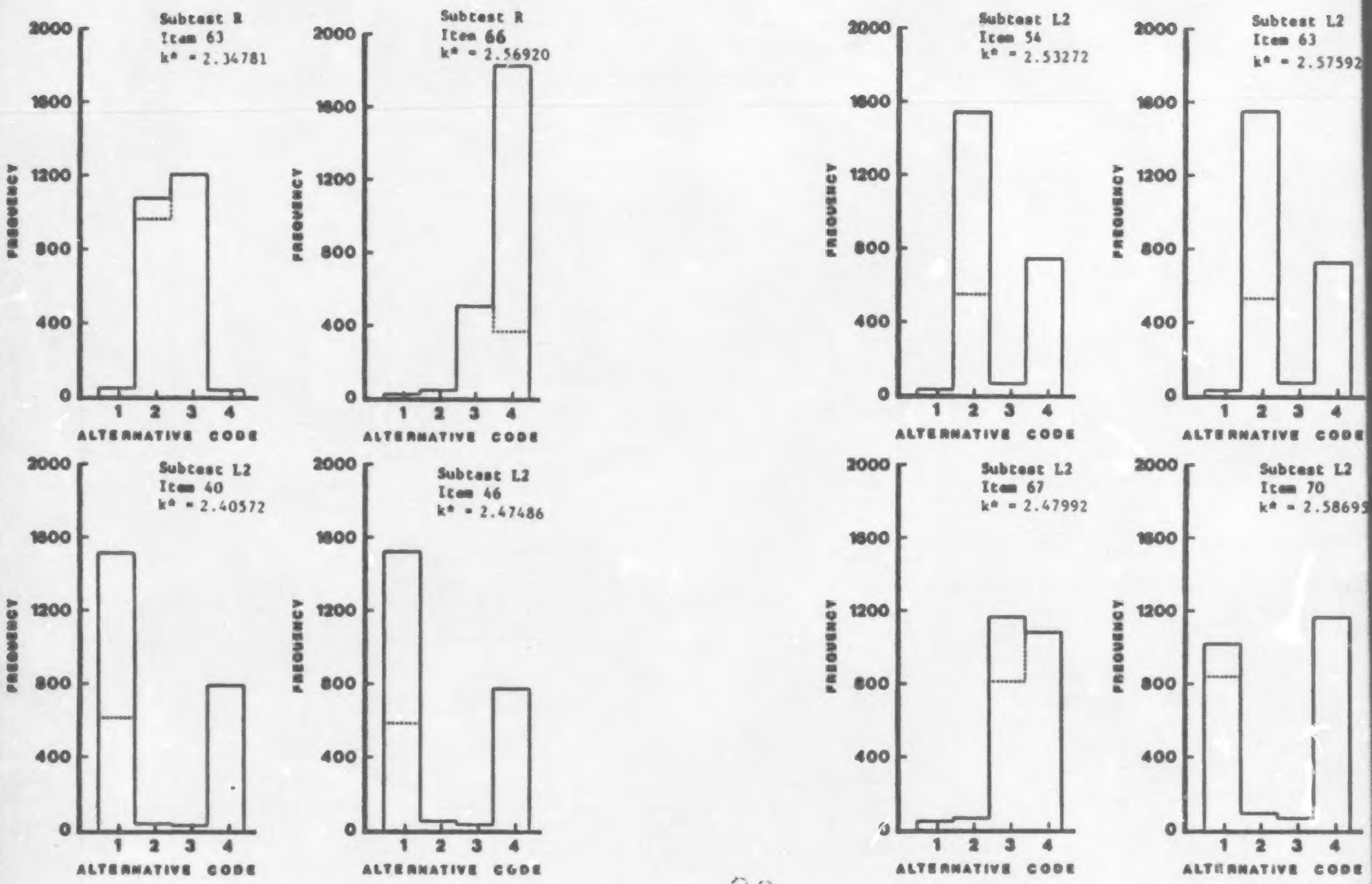


FIGURE 4-5 (Continued): Level 12.

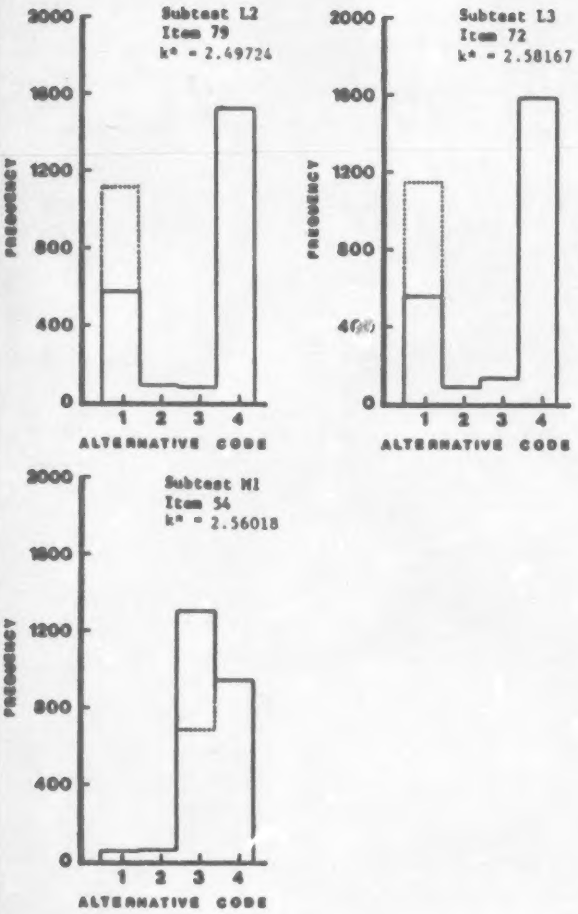


FIGURE 4-5 (Continued): Level 12.

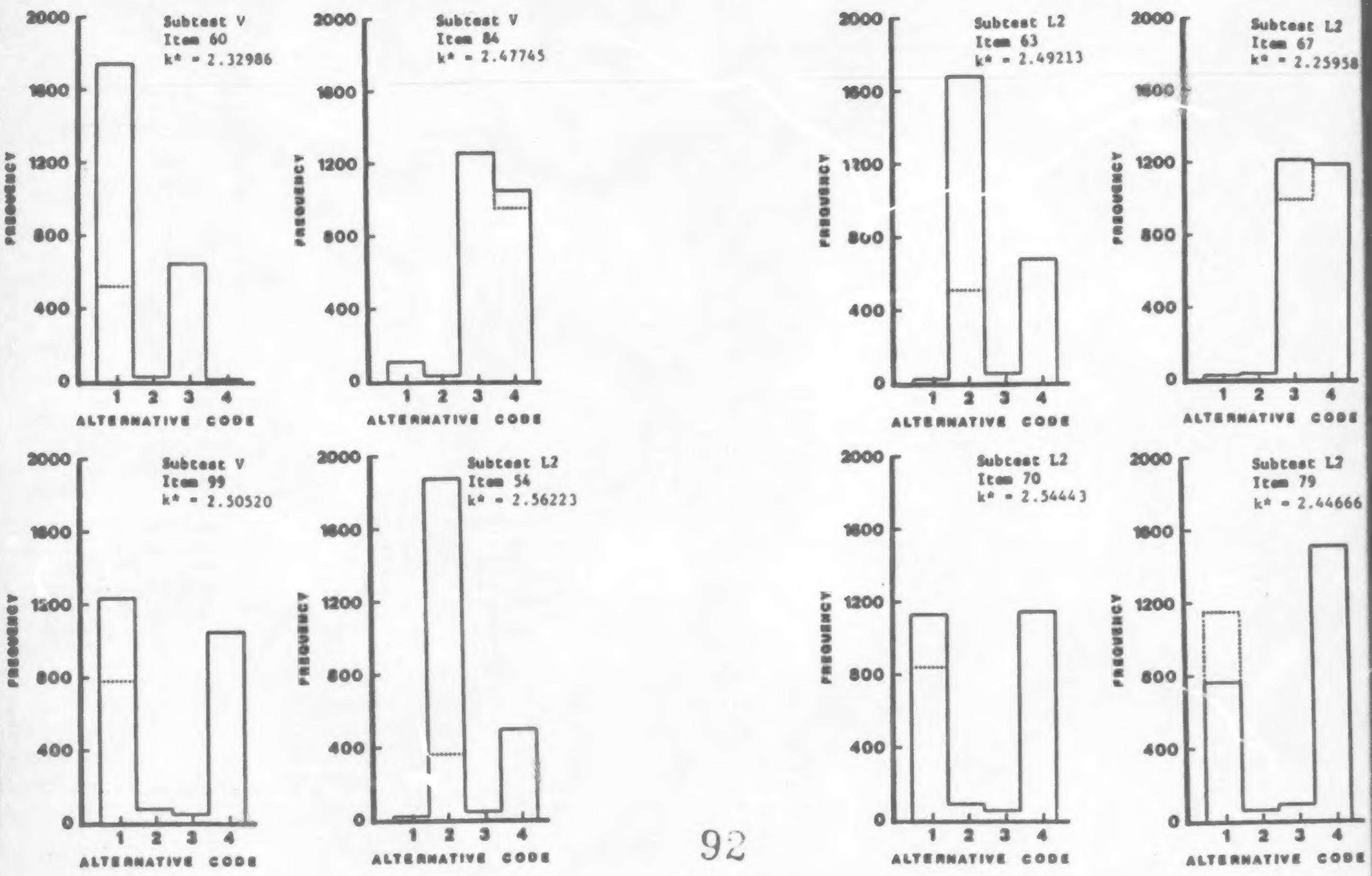


FIGURE 4-5 (Continued): Level 13.

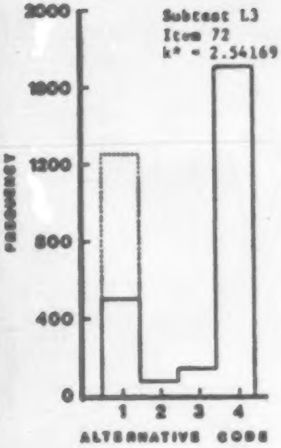


FIGURE 4-5 (Continued): Level 13.

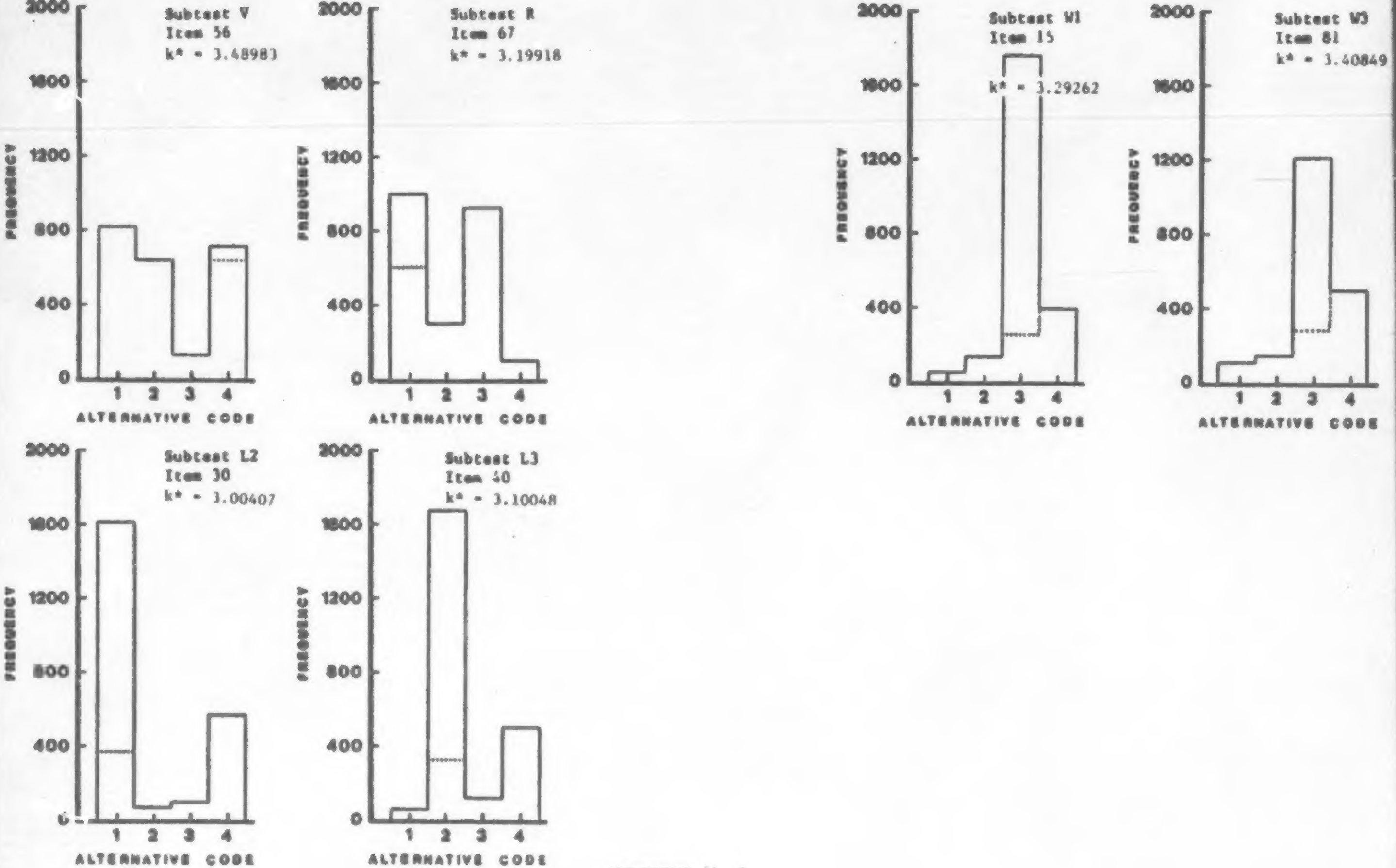


FIGURE 4-6

Frequency Distribution of the Examinees of Each of the Three Levels with Respect to Their Responses to Each Test Item Whose Index k* Is Greater than, or Equal to, 3.0 and Less than 3.6 , with the Estimated Proportion of the Examinees Guessing Correctly (Dotted Line).

Level 11

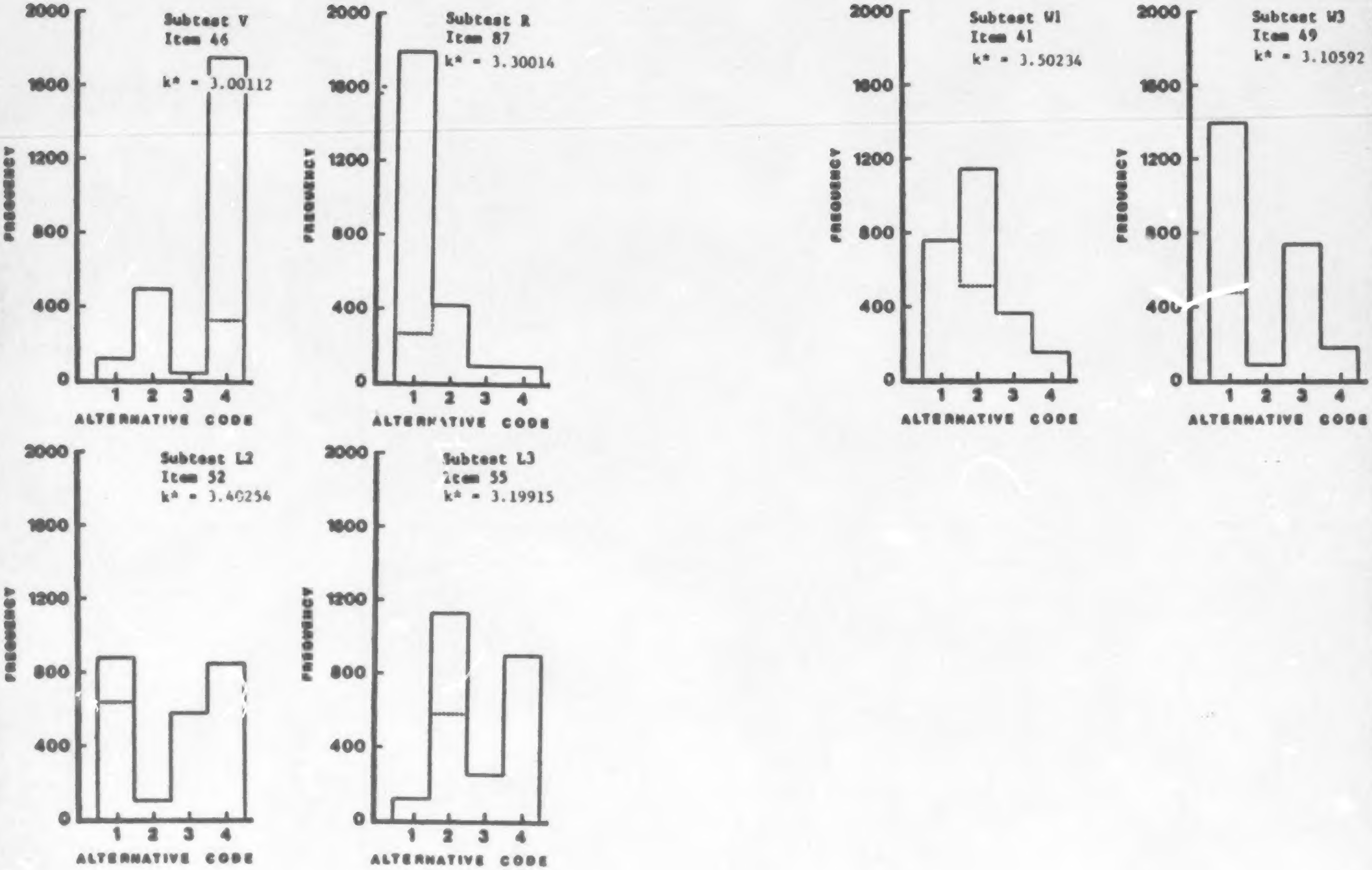


FIGURE 4-6 (Continued): Level 12.

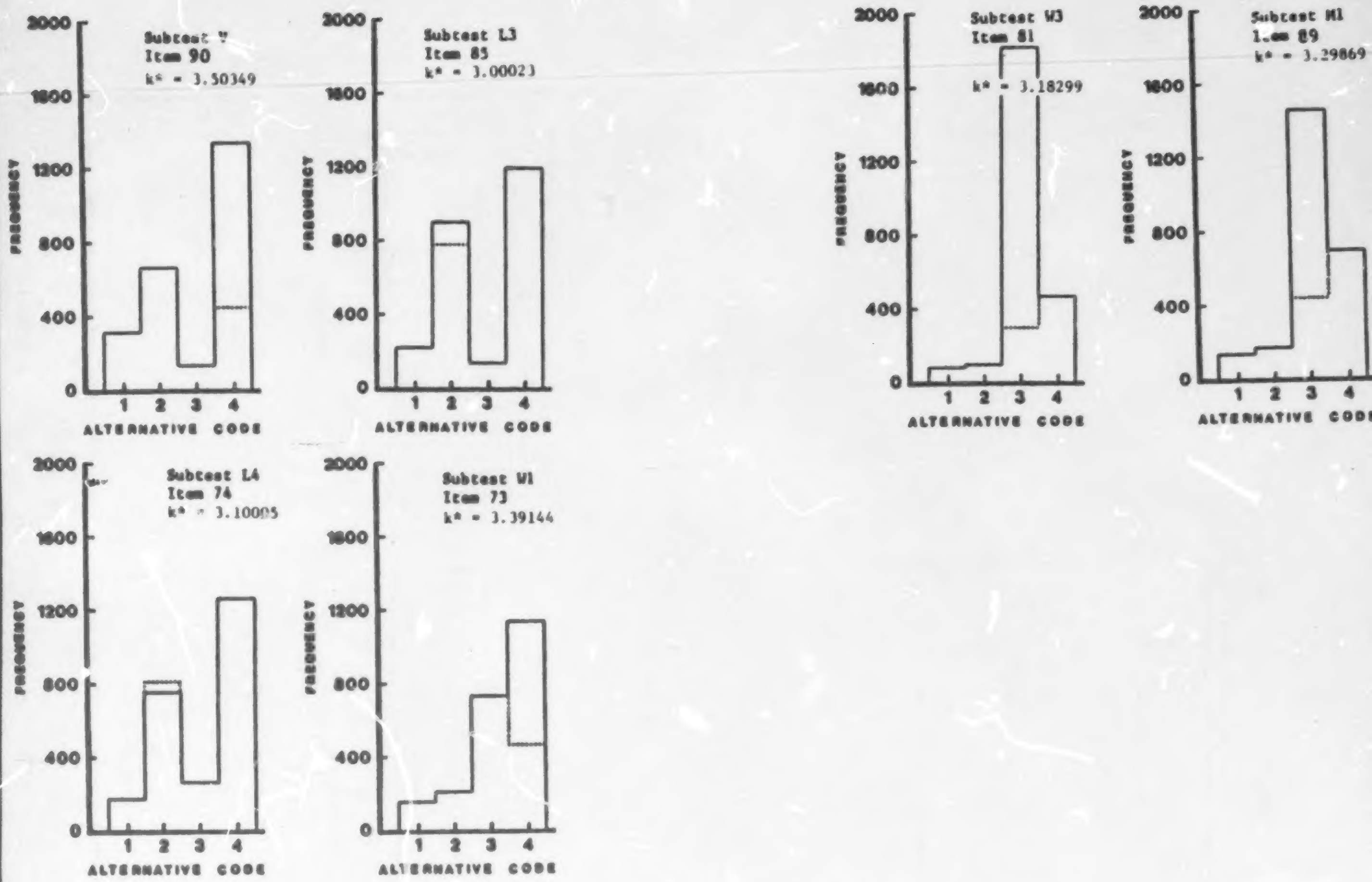
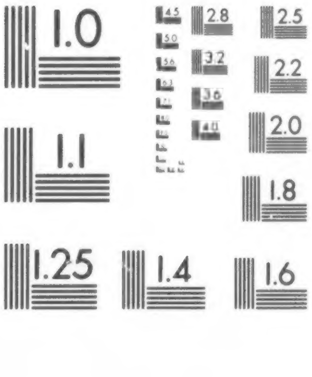


FIGURE 4-6 (Continued): Level 13.

2

F.R.


1.0
4.5
5.0
5.6
6.3
7.1
8.0
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

for the correct answer.

As an additional information, Figure 4-7 presents similar histograms for W1-13 and R-26 of Level 11, since their probabilities obtained by the chi-square test, which was conducted upon the original data, turned out to be 0.005 and 0.001, respectively, although their values of estimated Index k* are less than 3.9, i.e., 3.87789 for W1-13 and 3.83109 for R-26 .

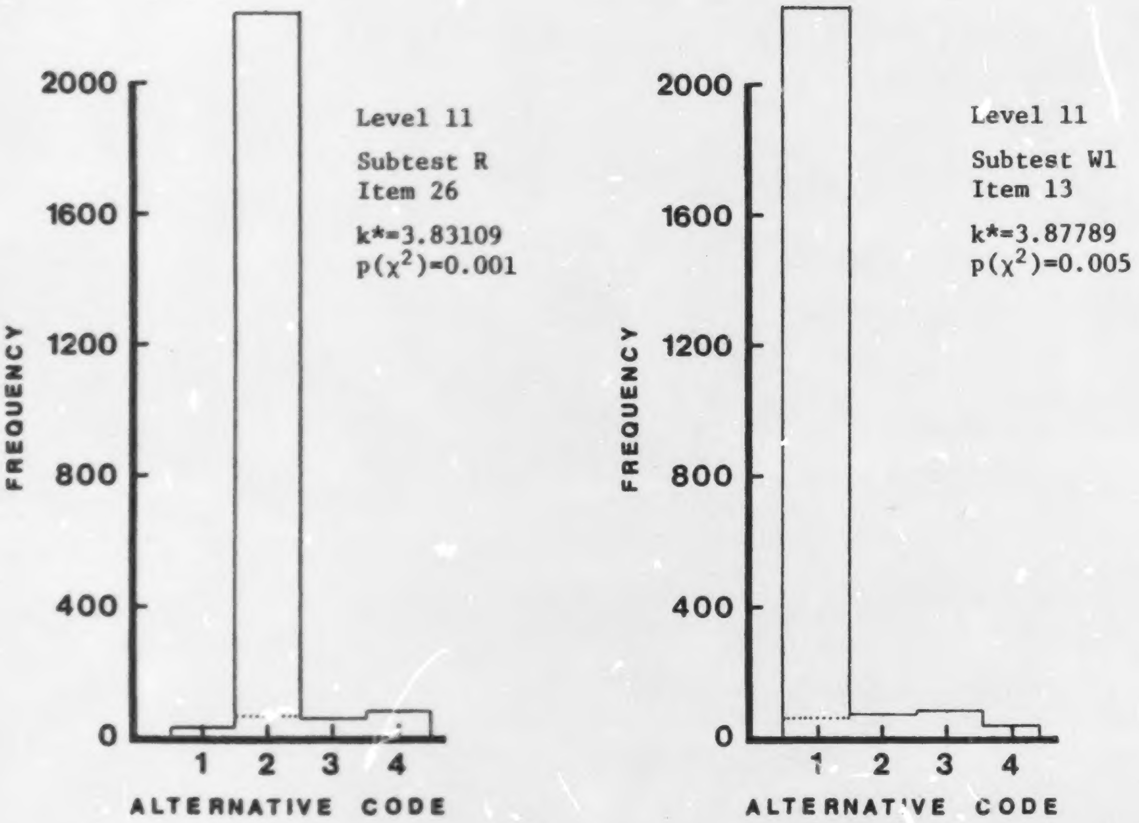


FIGURE 4-7

Frequency Distribution of Examinees with Respect to Their Responses to Each Test Item Whose Index k* Is Less Than 3.9 (Revised Data) and Whose Probability Obtained by the Chi-Square Test (Original Data) is 0.001 or Greater, with the Estimated Proportion of the Examinees Guessing Correctly (Dotted Line).

V Comparisons of the Results of Test Items Which Were Administered to Levels 11, 12, and 13

There are certain test items which are included in all the three levels. They are items 58 through 66 of Subtest V, 80 through 98 of Subtest R, 58 through 66 of Subtest L1, 49 through 58 of Subtest L2, 49 through 58 of Subtest L3, 44 through 54 of Subtest L4, 38 through 47 of Subtest W1, 41 through 46 of Subtest W2, and 67 through 82 of Subtest W3. Thus there are nine items of Subtest V, which were administered to all the three groups of students, nineteen of Subtest R, nine of Subtest L1, ten of Subtest L2, ten of Subtest L3, eleven of Subtest L4, ten of Subtest W1, six of Subtest W2 and sixteen of Subtest W3, which make the total number of test items shared by all the three levels of test 100. There is no item which is included in all three levels for Subtests M1 and M2.

It is evident that, for the behavior of the test item to follow Equivalent Distractor Model, not only the value of estimated Index k* should be close to m for one level of examinees but also for all three levels. For this reason, it will be worthwhile to compare the results across the three levels for these one hundred test items which are included in all the three levels of test.

Tables 5-1 and 5-2 present the three values of estimated Index k* , which were rounded to the second digit after the decimal point, for each of the ninety-one four-alternative test items, and for each of the nine five-alternative test items, respectively. We can see that only seven four-alternative test items, i.e., V-61,

TABLE 5-1

Three Different Values of Index k* for Each of the Four Alternative Items That Were Administered to All Three Levels, with Their Respective Item Numbers.

Subtest V

Item Numbers		Level		
	(11)(12)(13)	11	12	13
58	35 19 1	3.95	3.92	3.73
59	36 20 2	3.69	3.38	2.88
60	37 21 3	2.99	2.68	2.33
61	38 22 4	3.99	3.97	3.99
62	39 23 5	3.74	3.68	3.53
63	40 24 6	3.94	3.91	3.82
64	41 25 7	3.24	3.03	2.79
65	42 26 8	3.86	3.72	3.57
66	43 27 9	3.82	3.86	3.90

TABLE 5-1 (Continued): Subtest R.

Item Numbers		Level		
	(11)(12)(13)	11	12	13
80	56 20 1	3.93	3.86	3.83
81	57 21 2	3.53	3.35	3.17
82	58 22 3	3.44	3.16	2.87
83	59 23 4	3.46	3.10	2.82
84	60 24 5	3.80	3.61	3.39
85	61 25 6	3.62	3.48	3.46
86	62 26 7	3.22	3.08	3.12
87	63 27 8	3.37	3.30	2.77
88	64 28 9	3.98	3.98	3.96
89	65 29 10	3.80	3.74	3.63
90	66 30 11	3.97	3.90	3.86
91	67 31 12	3.80	3.65	3.53
92	68 32 13	3.88	3.91	3.96
93	69 33 14	3.79	3.74	3.67
94	70 34 15	3.82	3.61	3.33
95	71 35 16	3.89	3.82	3.72
96	72 36 17	3.91	3.77	3.72
97	73 37 18	3.91	3.81	3.74
98	74 38 19	3.78	3.75	3.83

TABLE 5-1 (Continued): Subtest L2.

Item Numbers		Level		
	(11)(12)(13)	11	12	13
49	31 11 1	3.19	2.89	3.06
50	32 12 2	3.71	3.75	3.87
51	33 13 3	3.66	3.25	3.13
52	34 14 4	3.27	3.40	3.22
53	35 15 5	2.98	2.93	2.74
54	36 16 6	2.73	2.53	2.56
55	37 17 7	2.49	2.72	2.71
56	38 18 8	3.15	2.96	2.76
57	39 19 9	3.38	3.14	3.17
58	40 20 10	3.95	3.95	3.86

TABLE 5-1 (Continued): Subtest L3.

Item Numbers		Level		
	(11)(12)(13)	11	12	13
49	31 11 1	3.95	3.96	3.86
50	32 12 2	3.75	3.72	3.80
51	33 13 3	3.80	3.61	3.36
52	34 14 4	2.94	2.90	2.97
53	35 15 5	2.96	2.93	2.72
54	36 16 6	3.10	3.05	2.99
55	37 17 7	3.49	3.20	2.82
56	38 18 8	3.62	3.70	3.80
57	39 19 9	3.58	3.61	3.58
58	40 20 10	3.28	3.30	3.33

TABLE 5-1 (Continued): Subtest L4.

Item Numbers		Level		
	(11)(12)(13)	11	12	13
44	22 12 1	3.18	3.00	2.60
45	23 13 2	3.47	3.33	3.12
46	24 14 3	3.82	3.76	3.79
47	25 15 4	3.53	3.61	3.48
48	26 16 5	3.18	3.18	3.17
49	27 17 6	3.44	3.28	3.34
50	28 18 7	3.19	3.07	2.85
51	29 19 8	3.52	3.42	3.35
52	30 20 9	3.81	3.78	3.75
53	31 21 10	3.69	3.63	3.75
54	32 22 11	3.35	3.41	3.39

TABLE 5-1 (Continued): Subtest W1.

Item Numbers		Level		
	(11)(12)(13)	11	12	13
38	27 11 1	3.91	3.82	3.76
39	28 12 2	3.65	3.64	3.37
40	29 13 3	3.92	3.89	3.83
41	30 14 4	3.71	3.50	3.34
42	31 15 5	3.57	3.59	3.67
43	32 16 6	3.96	3.96	3.89
44	33 17 7	3.84	3.56	3.60
45	34 18 8	3.96	3.95	3.94
46	35 19 9	3.95	3.94	3.91
47	36 20 10	3.94	3.85	3.87

TABLE 5-1 (Continued): Subtest W2.

Item Numbers		Level		
	(11)(12)(13)	11	12	13
41	21 9 1	3.97	3.87	3.93
42	22 10 2	3.81	3.76	3.67
43	23 11 3	3.94	3.87	3.72
44	24 12 4	3.91	3.99	4.00
45	25 13 5	3.95	3.97	3.96
46	26 14 6	3.63	3.64	3.75

TABLE 5-1 (Continued): Subtest W3.

Item Numbers		Level		
	(11)(12)(13)	11	12	13
67	41 25 1	3.34	3.07	2.78
68	42 26 2	3.00	2.75	2.68
69	43 27 3	3.92	3.77	3.67
70	44 28 4	3.93	3.92	3.97
71	45 29 5	3.34	3.38	3.33
72	46 30 6	3.25	2.99	3.10
73	47 31 7	3.80	3.56	3.57
74	48 32 8	3.79	3.78	3.75
75	49 33 9	3.37	3.13	3.22
76	50 34 10	3.20	3.32	3.24
77	51 35 11	3.63	3.38	3.36
78	52 36 12	3.92	3.76	3.83
79	53 37 13	3.71	3.63	3.57
80	54 38 14	3.49	3.46	3.53
81	55 39 15	3.41	3.06	3.18
82	56 40 16	3.71	3.54	3.44

TABLE 5-2

Three Different Values of Index k* for Each of the Five Alternative Items That Were Administered to All Three Levels, with Their Respective Item Numbers.

Subtest L1

Item Numbers		Level		
	(11)(12)(13)	11	12	13
58	35 19 1	2.53	2.36	2.39
59	36 20 2	3.54	3.03	2.92
60	37 21 3	4.17	3.99	3.87
61	38 22 4	3.79	3.62	3.33
62	39 23 5	2.90	2.83	2.67
63	40 24 6	4.31	4.14	4.00
64	41 25 7	3.46	3.28	3.20
65	42 26 8	3.66	3.36	3.31
66	43 27 9	4.36	4.32	3.94

R-88, W1-45, W1-46, W2-44, W2-45 and W3-70, have three estimates of Index k* all of which are greater than, or equal to, 3.9 . If we shift this critical value from 3.9 to 3.8 , these seven four-alternative test items are joined by eleven more items, i.e., V-63, V-66, R-80, R-90, R-92, L2-58, L3-49, W1-40, W1-43, W1-47 and W2-41 . There are no five-alternative test items of Subtest L1 which are comparable to these eighteen four-alternative test items.

We have selected Subtests R and W3, which have the two largest numbers of shared test items, as our examples, and drawn the set of three histograms, which are similar in nature to those in Figures 4-4, 4-5 and 4-6, for the set of three frequency distributions for each item. Figures 5-1 and 5-2 present the resultant nineteen sets of histograms for Subtest R, and the sixteen sets for Subtest W3, respectively.

It is interesting to note that some items show evidence of differential information provided by separate wrong answers. For example, alternative 4 of R-80 seems to attract students of intermediate reading ability, while alternative 1 of the same item appears to attract students of lower levels of ability. It is clear that many items have one or more effective distractors, and, among others, alternative 2 of R-86 proved to be powerful. Most histograms have some regularities in the way the frequencies change across the three levels, which suggest that the examinees selected their answers intentionally rather than by random guessing.

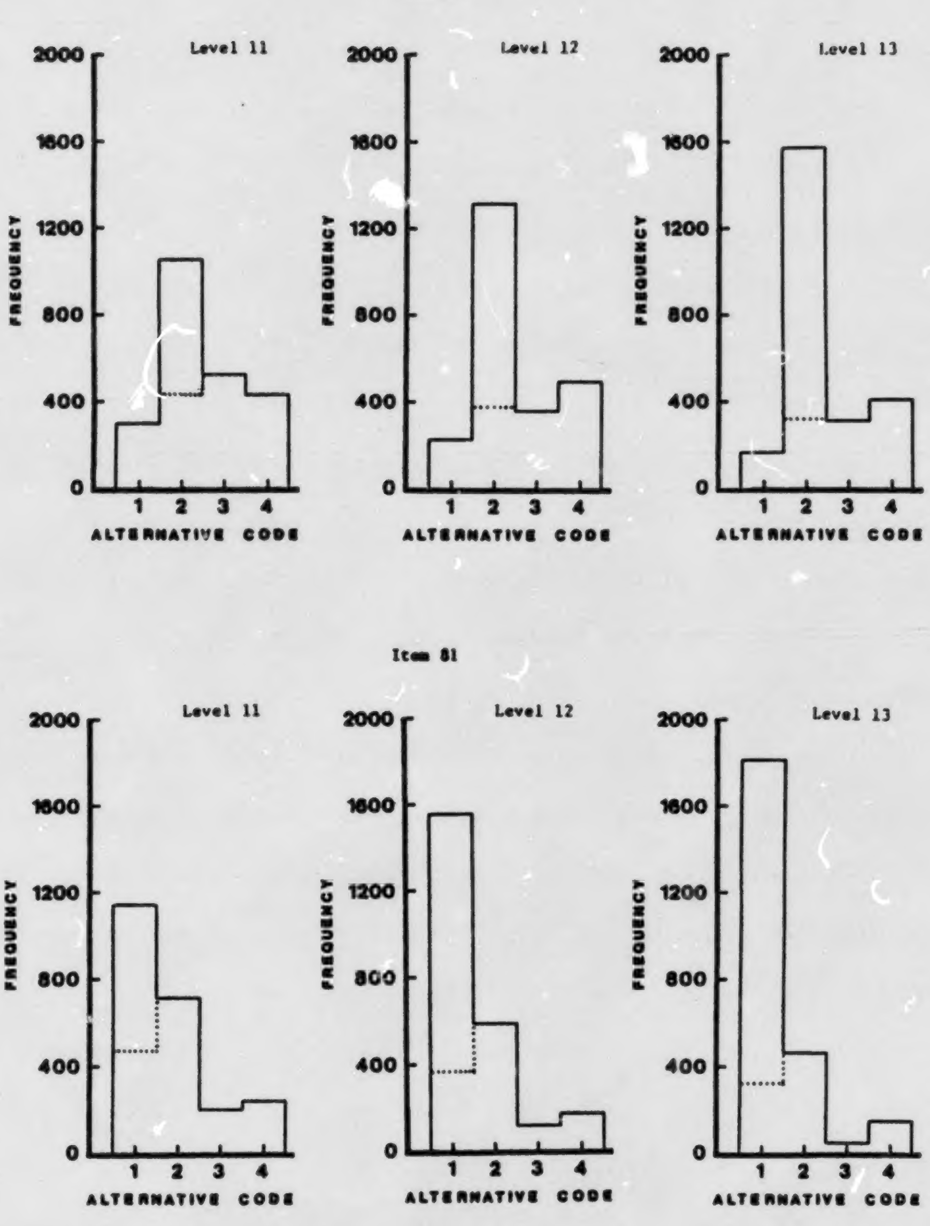
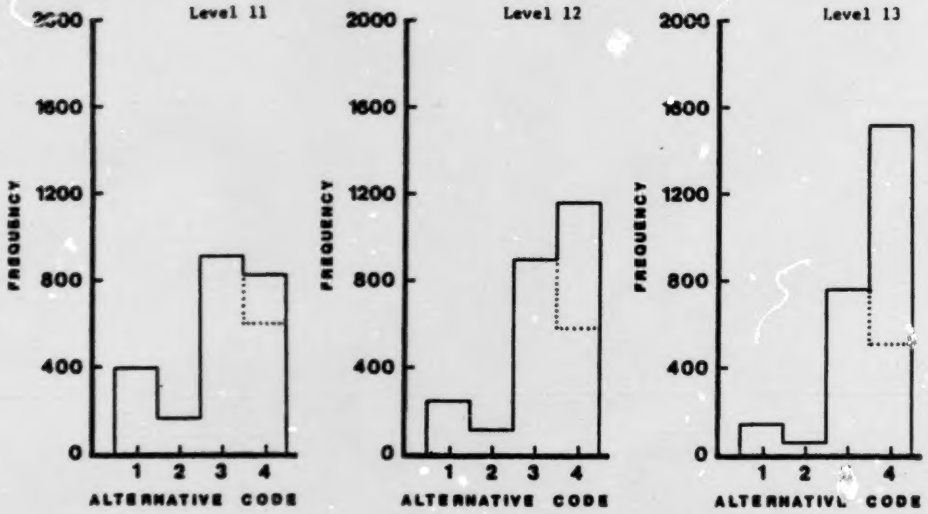


FIGURE 5-1

Comparison of the Three Frequency Distributions of Examinees with Respect to Their Choices of Alternatives for Each of the Twenty Items of Subtest R, Which Were Administered to All Three Levels of Students, with the Estimated Proportion of Examinees Guessing Correctly (Dotted Line).

Item 82



Item 83

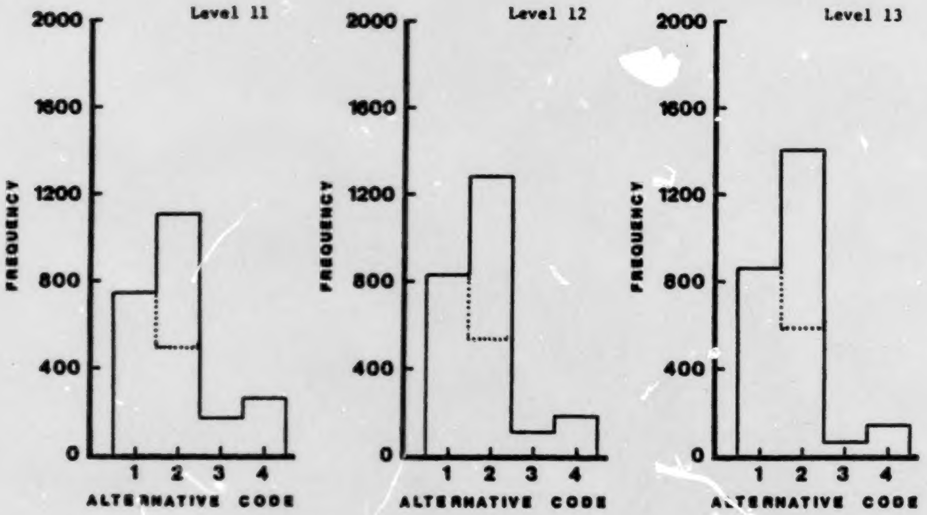
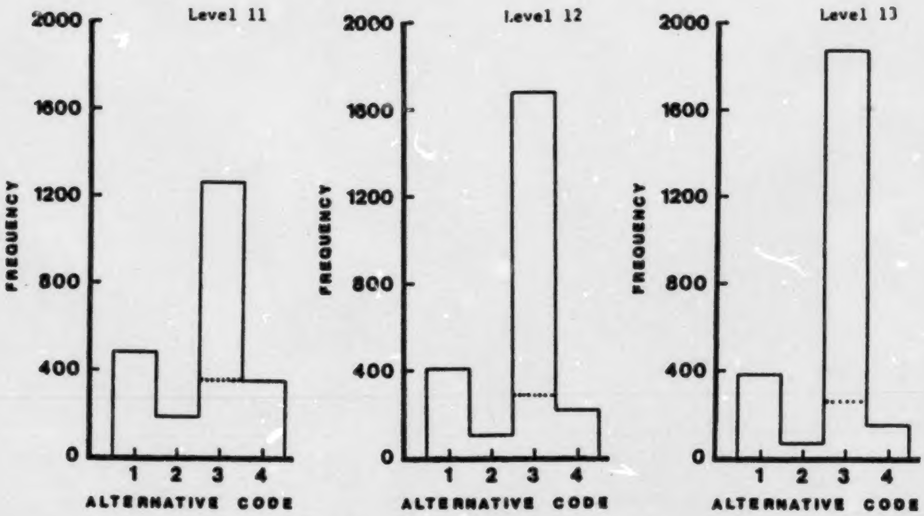


FIGURE 5-1 (Continued): Subtest R.

Item 84



Item 85

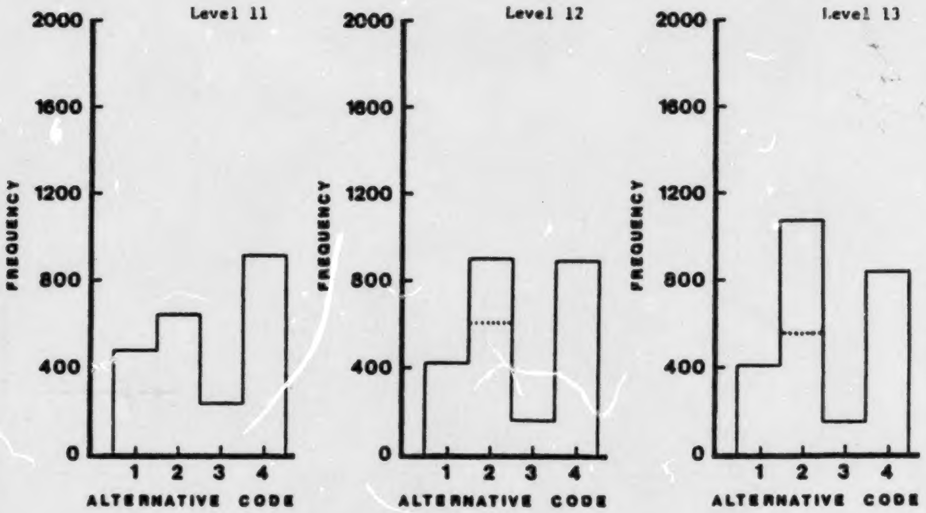
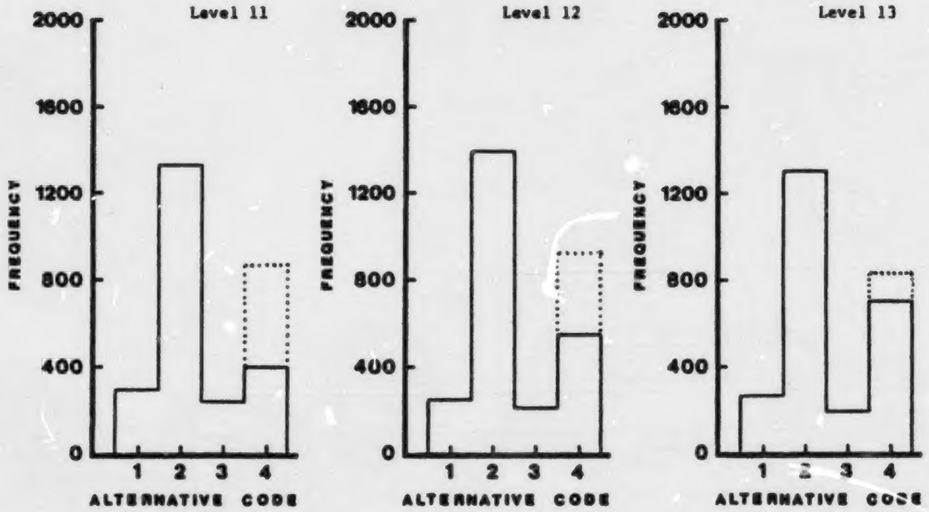


FIGURE 5-1 (Continued): Subtest R.

Item 86



Item 87

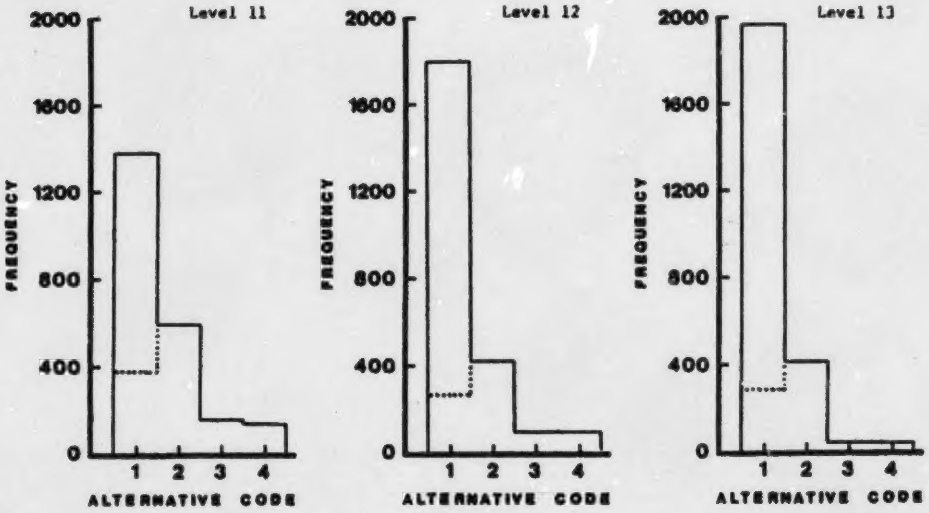
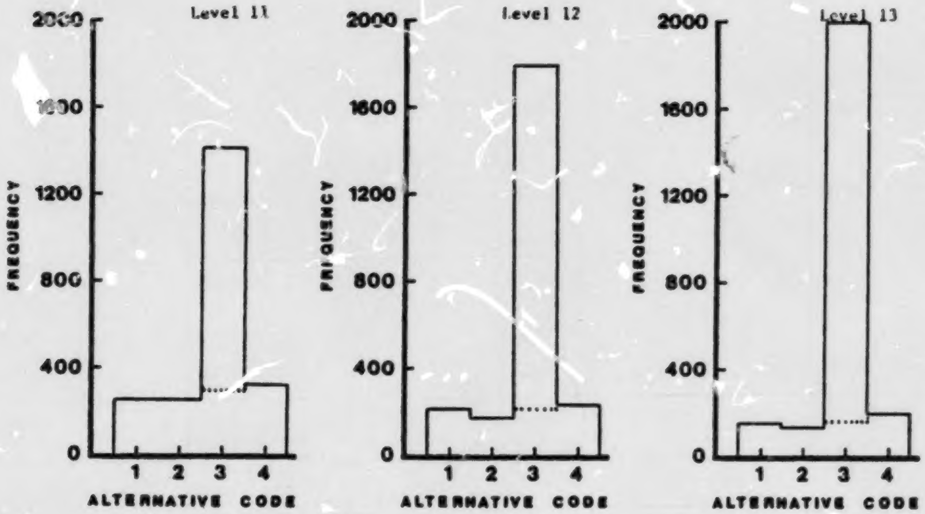


FIGURE 5-1 (Continued): Subtest R.

Item 88



Item 89

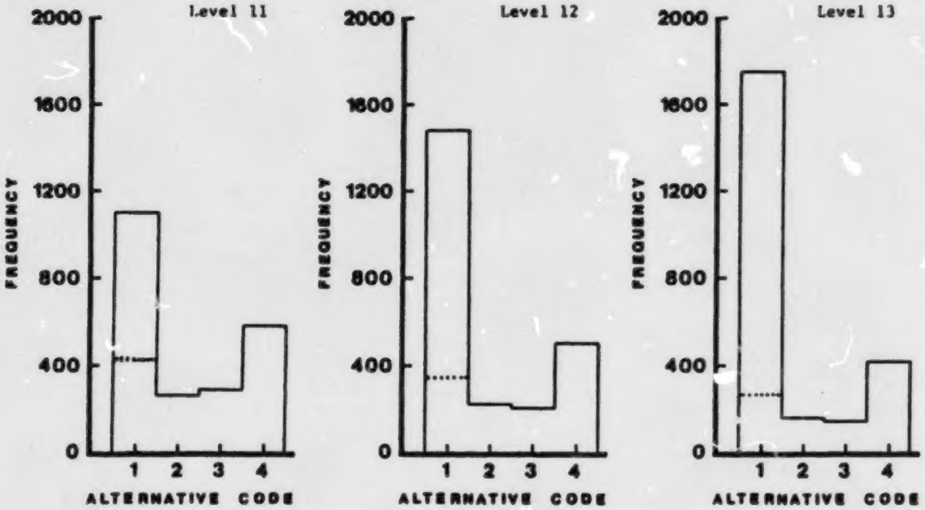
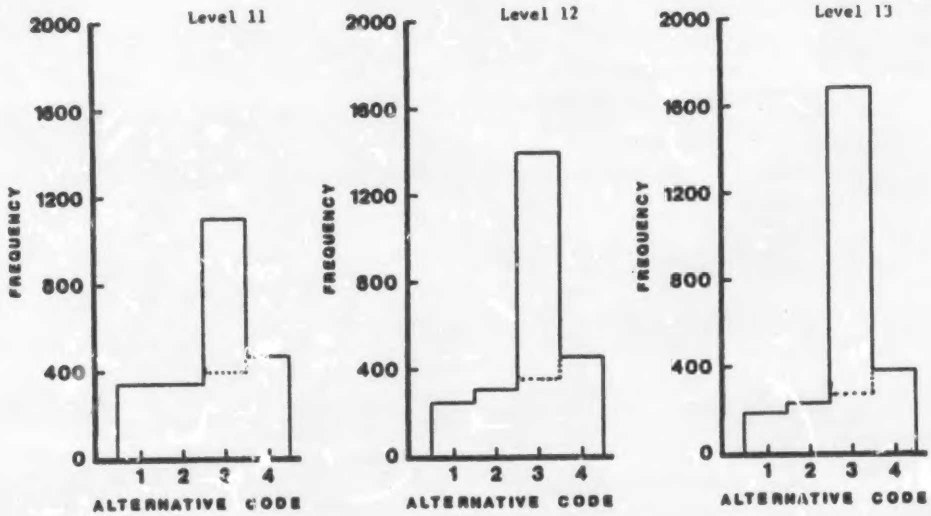


FIGURE 5-1 (Continued): Subtest R.

Item 90



Item 91

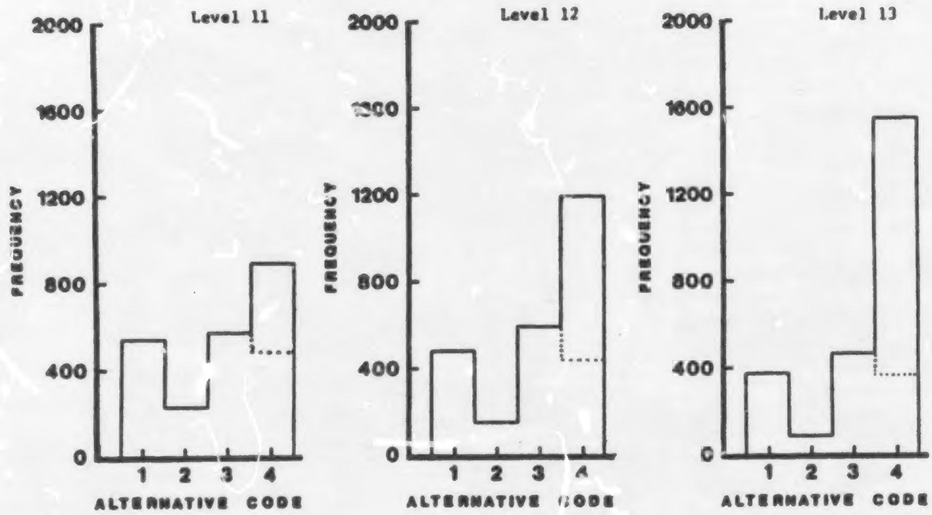


FIGURE 5-1 (Continued): Subtest R.

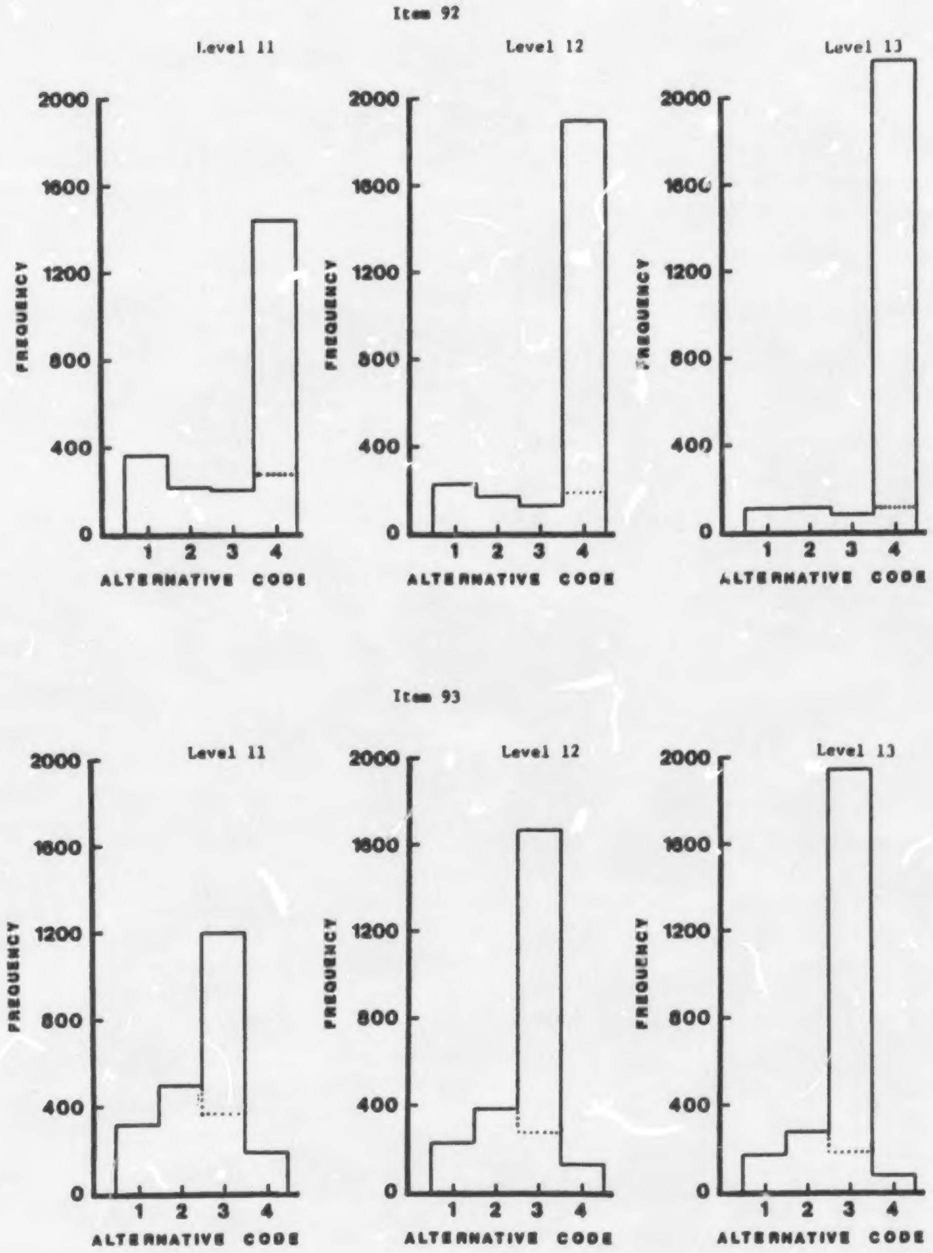
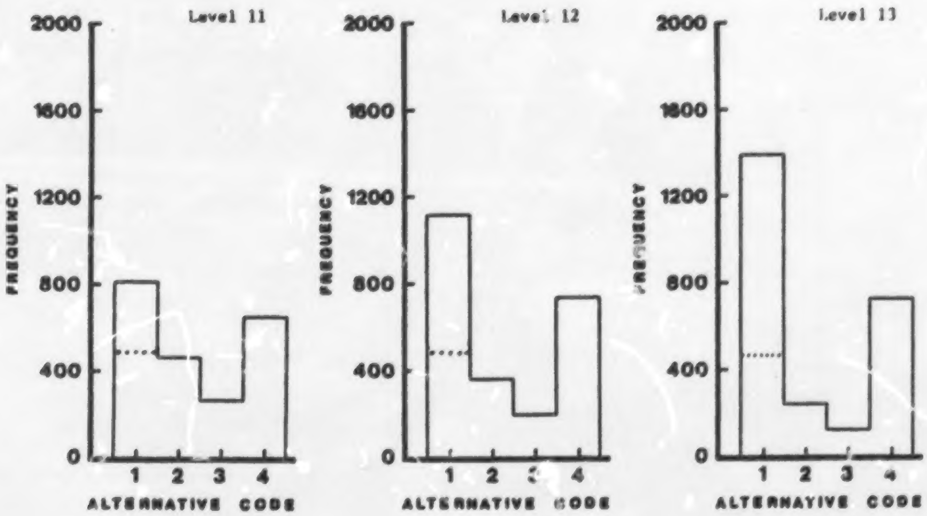


FIGURE 5-1 (Continued): Subtest R.

Item 94



Item 95

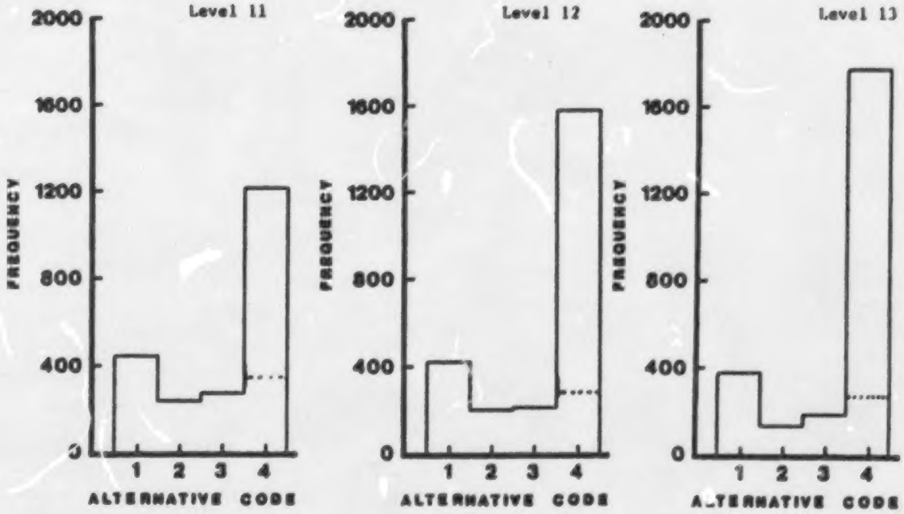


FIGURE 5-1 (Continued): Subtest R.

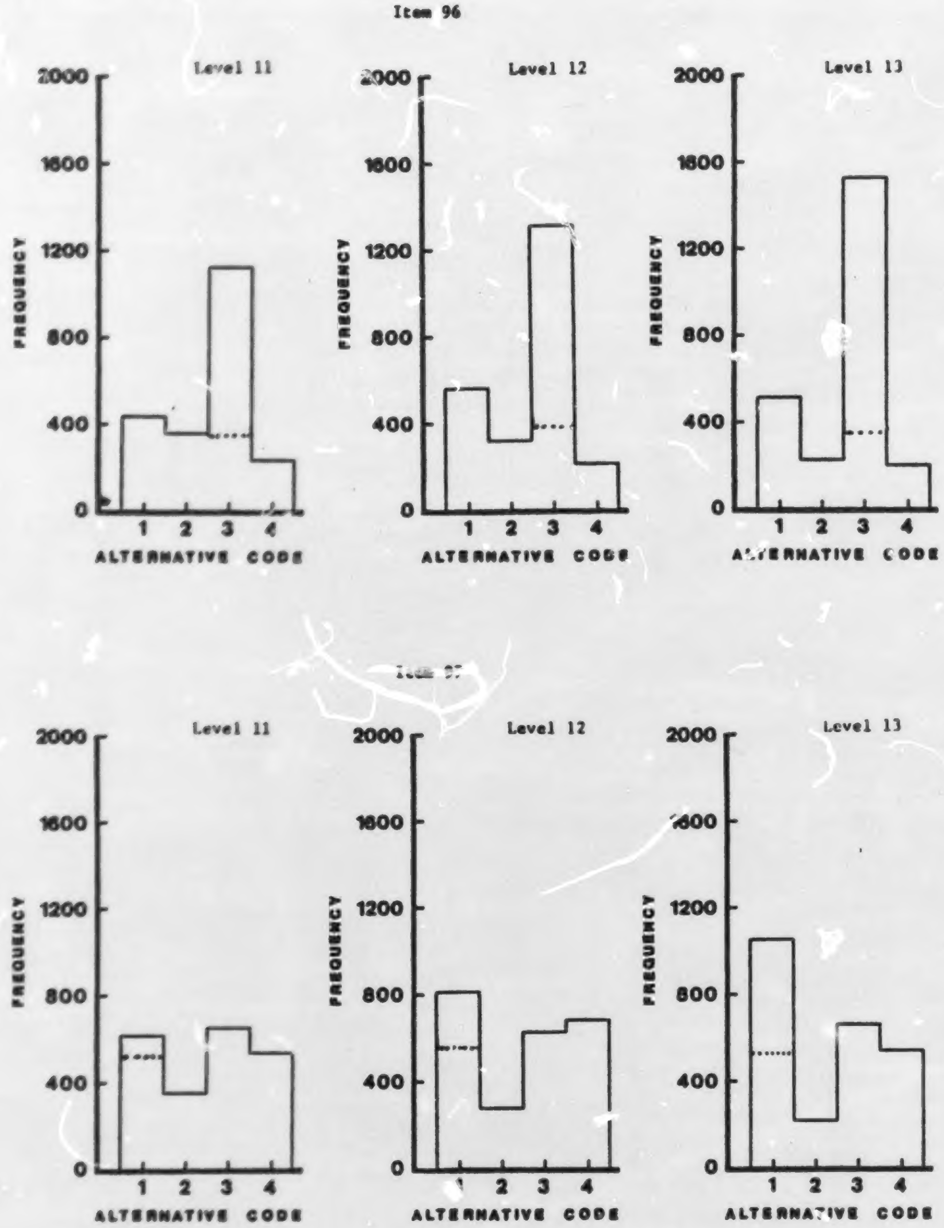


FIGURE 5-1 (Continued): Subtest R.

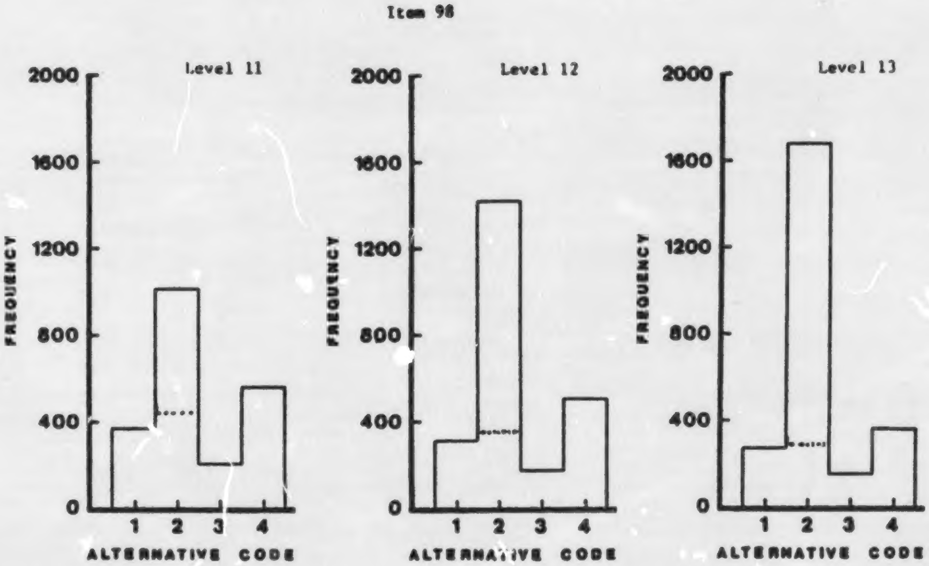


FIGURE 5-1 (Continued): Subtest R.

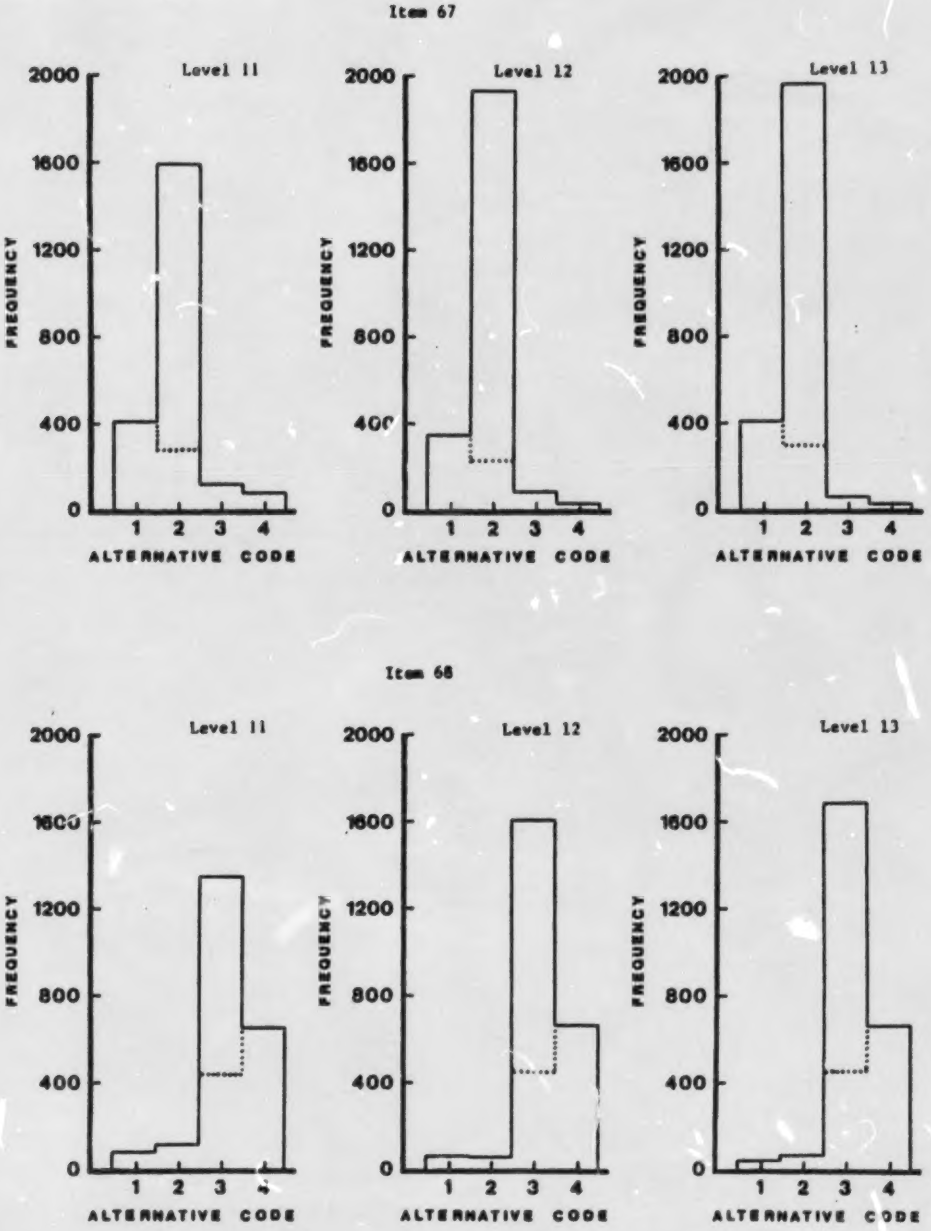
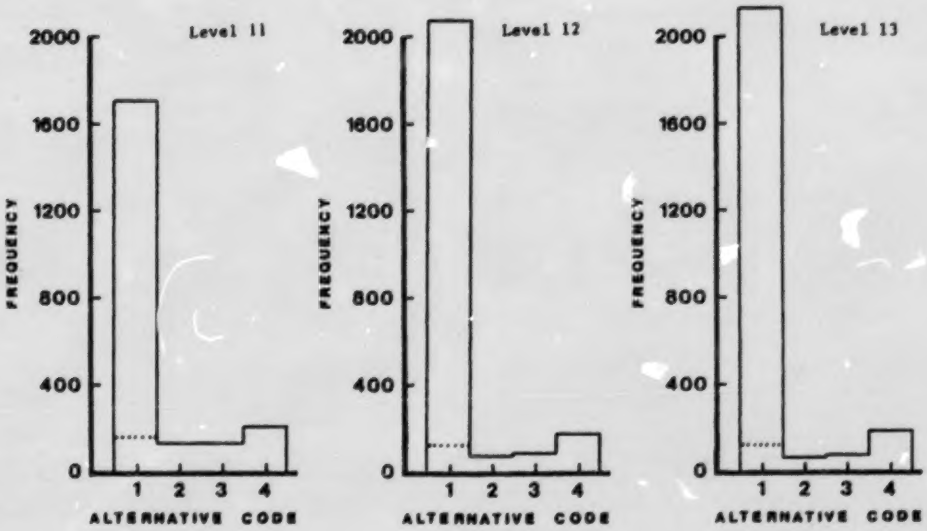


FIGURE 5-2

Comparison of the Three Frequency Distributions of Examinees with Respect to Their Choices of Alternatives for Each of the Sixteen Items of Subtest W3, Which Were Administered to All Three Levels of Students, with the Estimated Proportion of Examinees Guessing Correctly (Dotted Line).

Item 69



Item 70

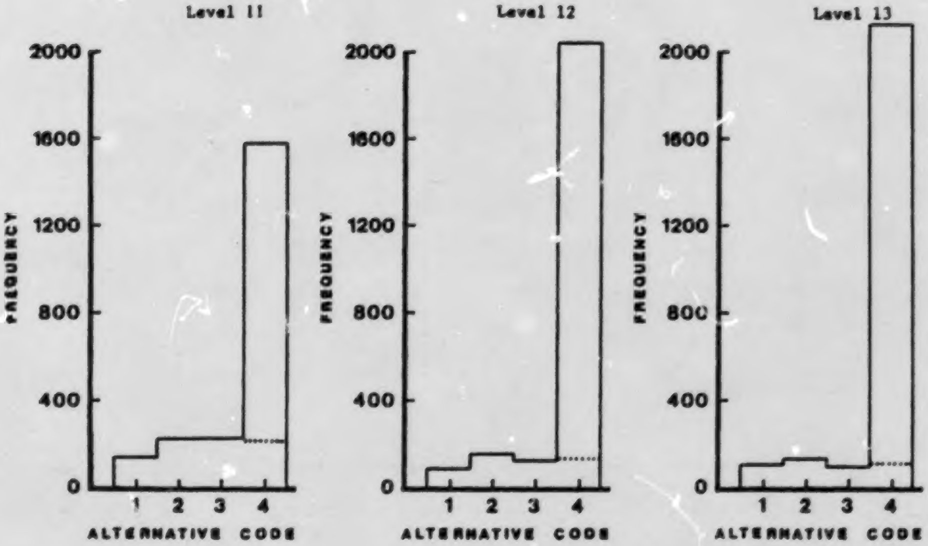
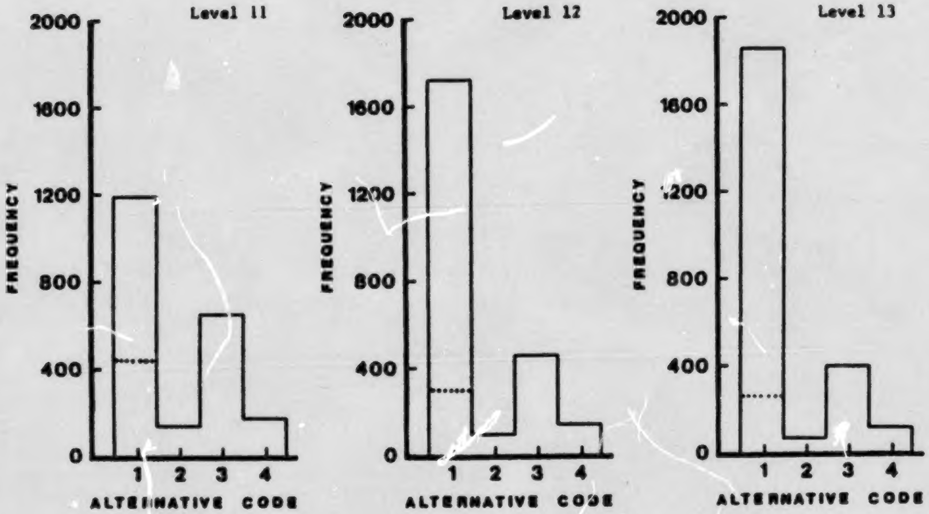


FIGURE 5-2 (Continued): Subtest W3.

Item 71



Item 72

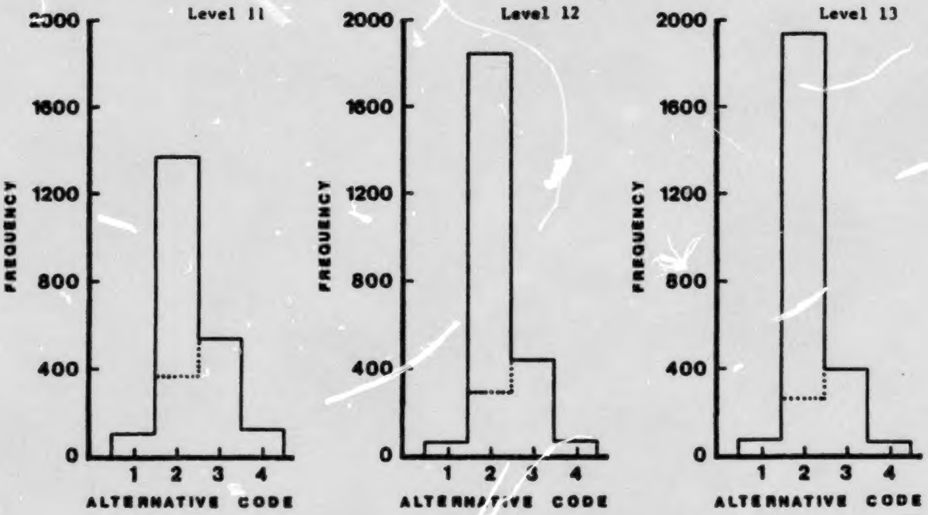
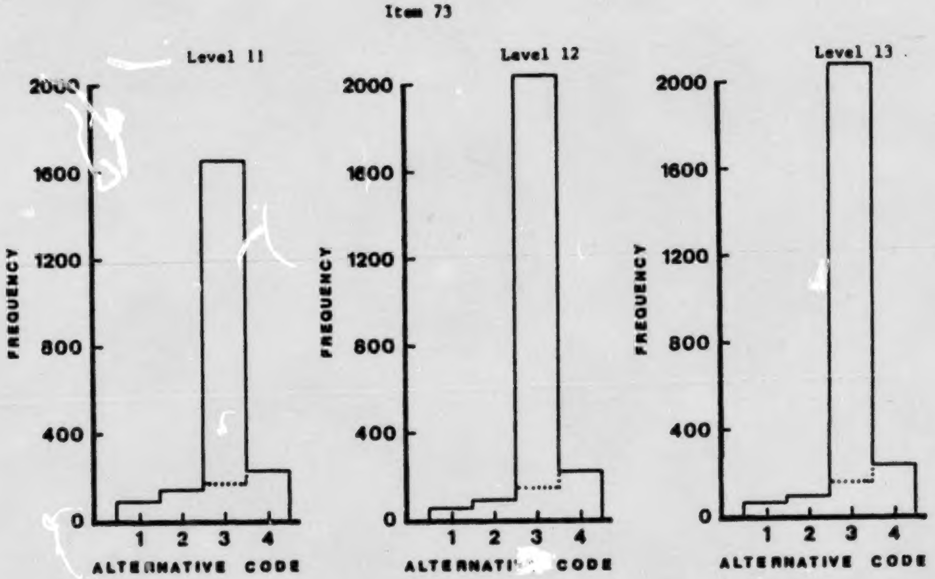


FIGURE 5-2 (Continued): Subtest W3.



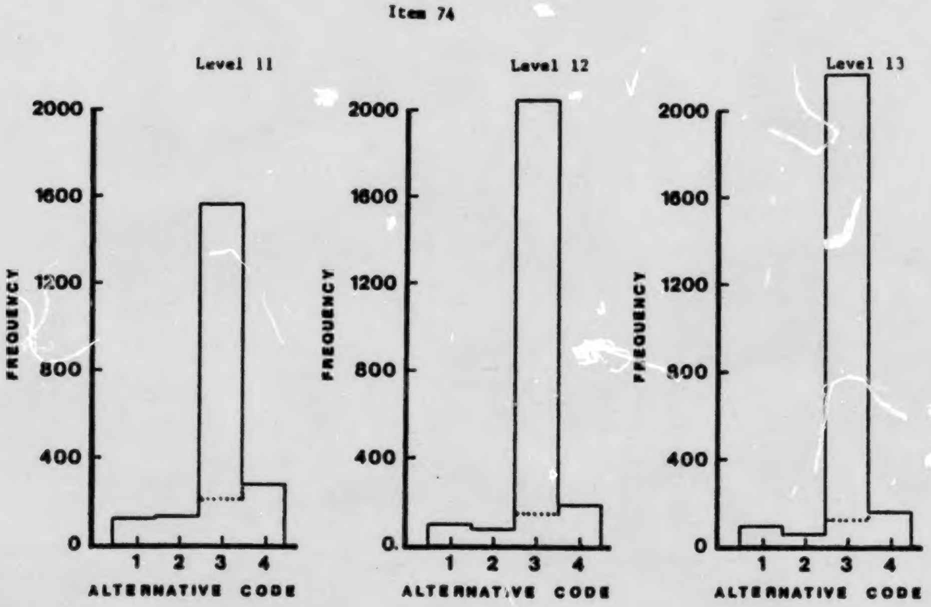
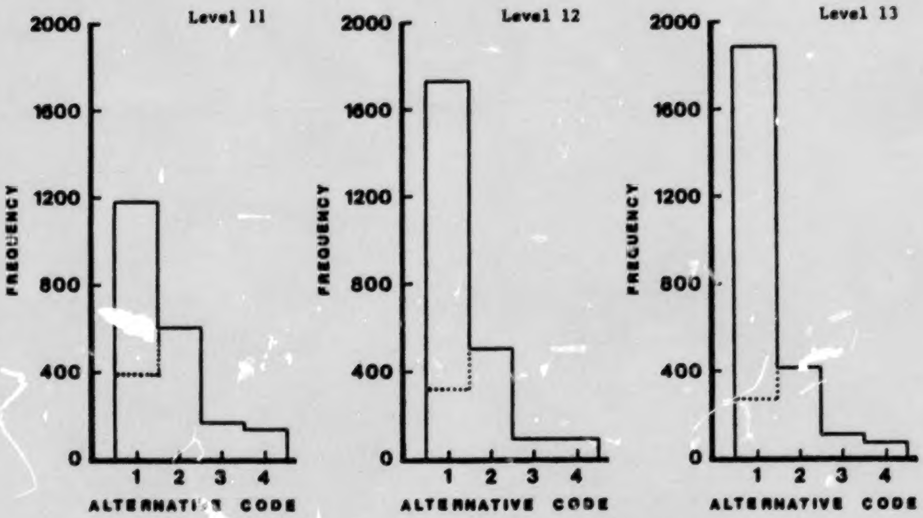


FIGURE 5-2 (Continued): Subtest W3.

Item 75



Item 76

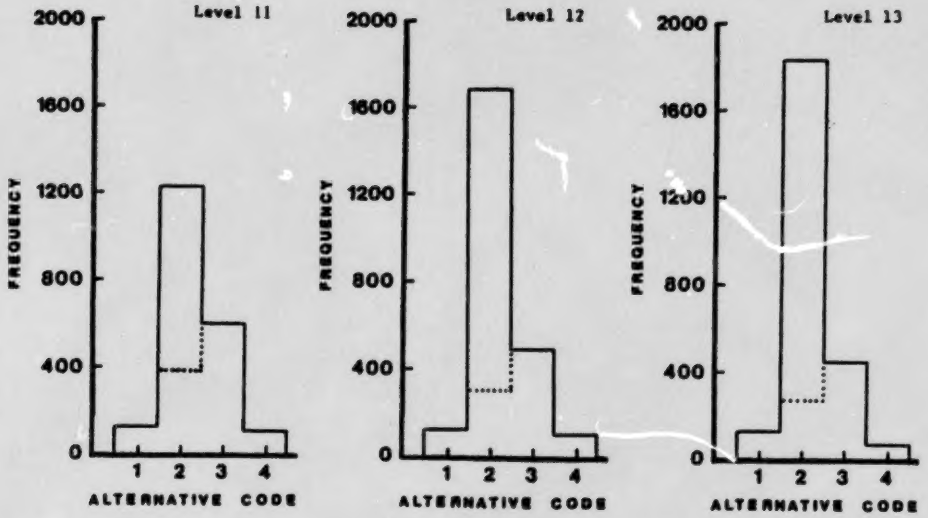
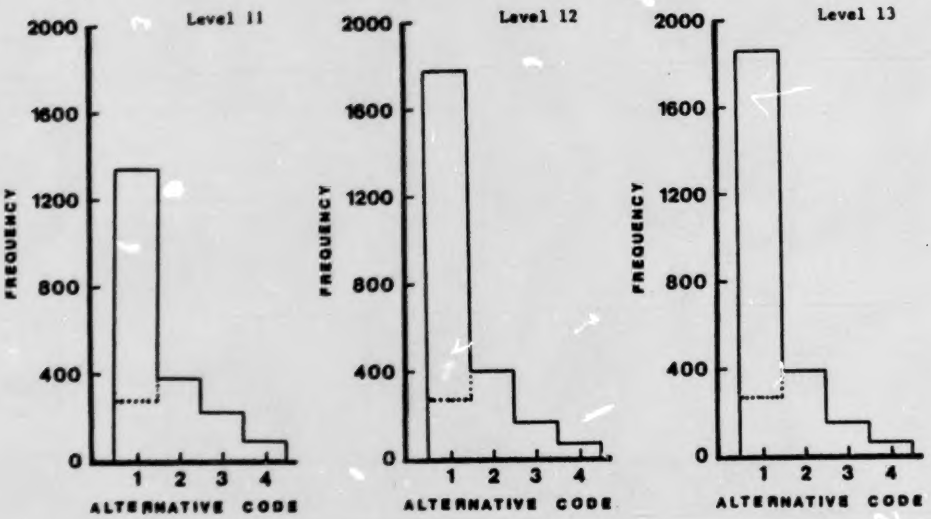


FIGURE 5-2 (Continued): Subtest W3.

Item 77



Item 78

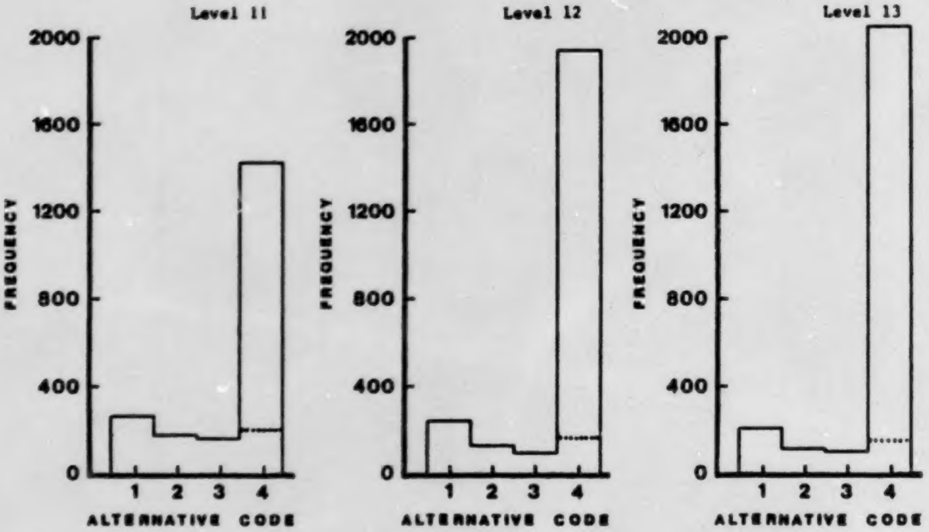
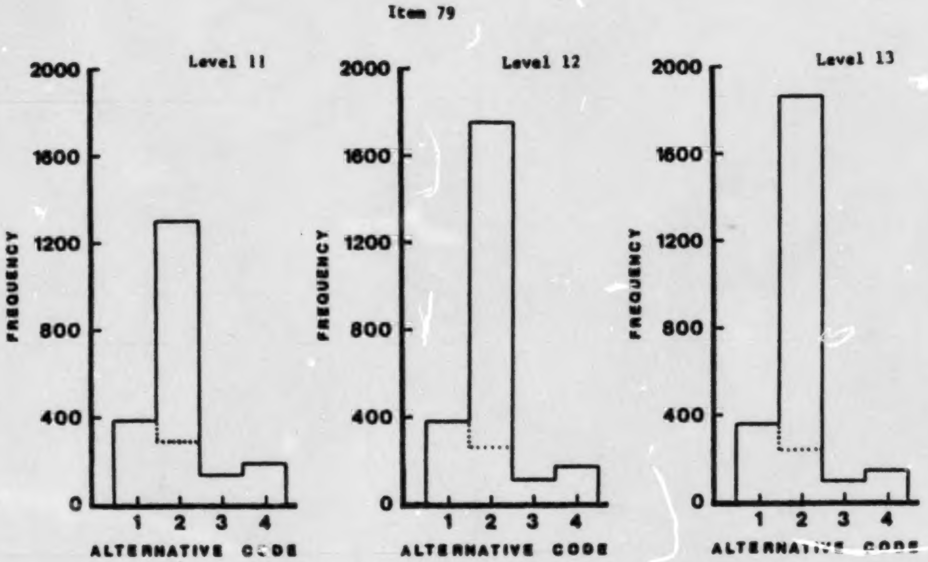


FIGURE 5-2 (Continued): Subtest W3.



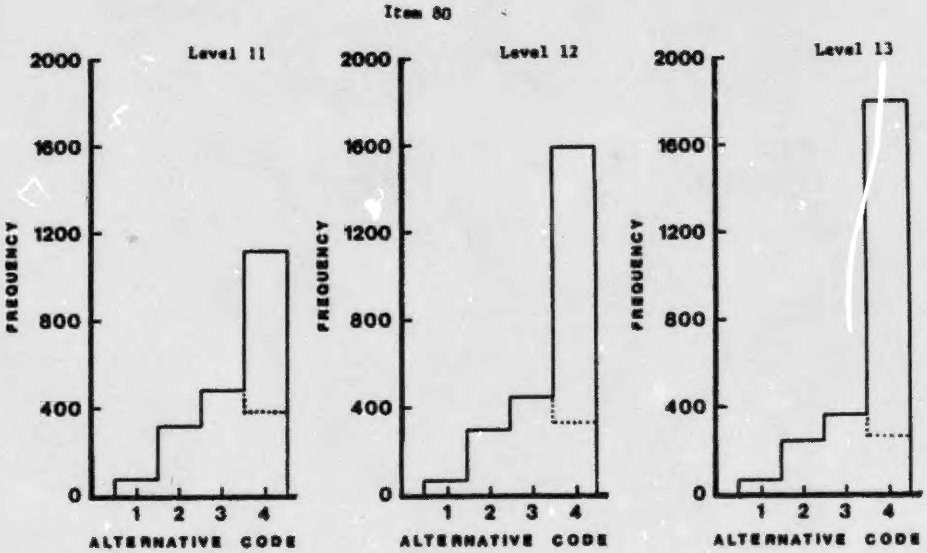
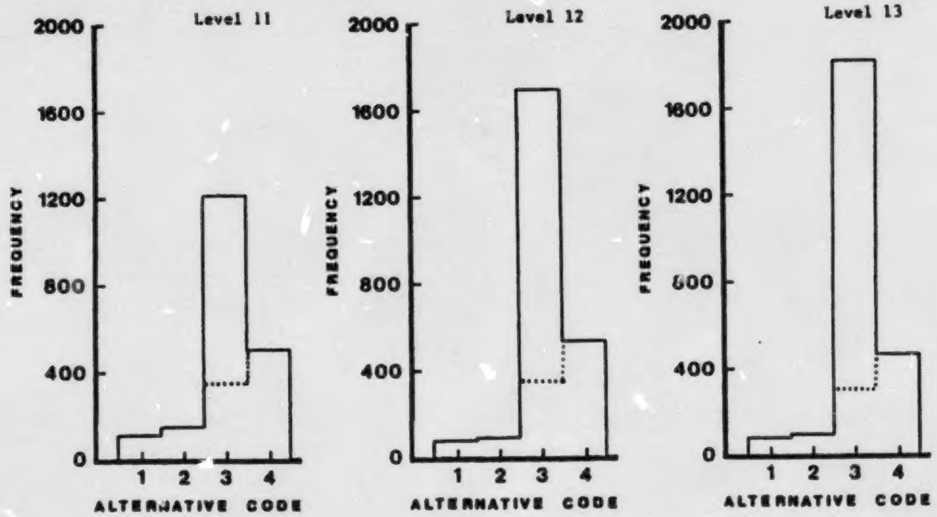


FIGURE 5-2 (Continued): Subtest W3.

Item 81



Item 82

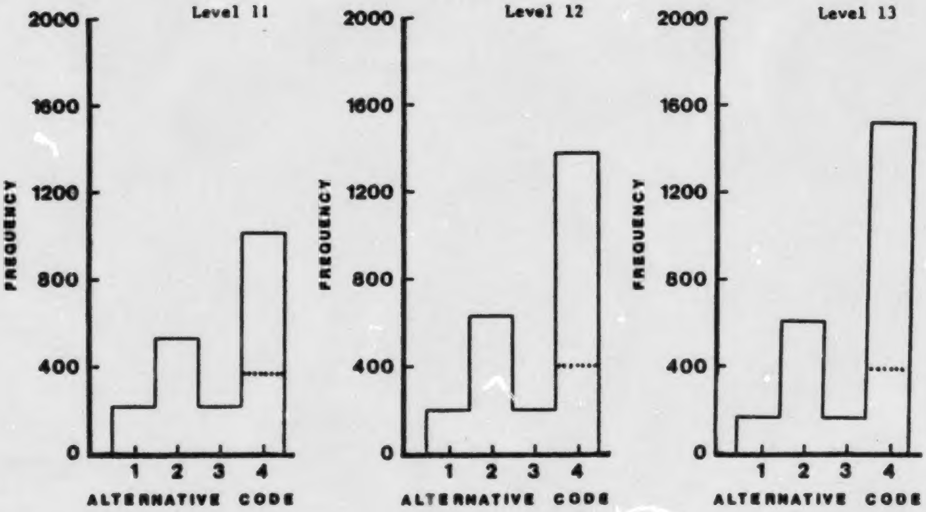


FIGURE 5-2 (Continued): Subtest W3.

VI Discussion and Conclusions

We have seen in the previous chapters that, generally speaking, evidence indicates the direction of Informative Distractor Model, rather than that of Equivalent Distractor Model. This is especially true with the test items for measuring language skills, i.e., those of Subtests L1, L2, L3 and L4. There are certain test items, however, for which the direction of Equivalent Distractor Model may be more suitable, as we have discussed in the preceding two chapters. For these reasons, it will be wise to take the strategy of adopting theory and method which make it possible for us to handle the test items following either general direction.

In so doing, perhaps the most promising way is to adopt the methods and approaches for estimating the operating characteristics without assuming any mathematical form. The selection of items which can be used as a substitute for the Old Test will be the most crucial point in that process. If it succeeds, then we shall be able to conduct the item analysis in the true sense of the word, which leads to the evaluation and suggestions for improvement of each test item.

REFERENCES

[1] Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord and M.R. Novick; *Statistical theories of mental test scores*. Addison-Wesley, 1968, Chapters 17-20.

[2] Hieronymous, A. N. and E. F. Lindquist. *Iowa test of basic skills* (Levels ed.), Form 6. Boston: Houghton-Mifflin Company, 1971.

[3] Iowa Basic Skills Testing Program. *Iowa test of basic skills teacher's manual*. Iowa City, Iowa: University of Iowa, 1971.

[4] Lord, F.M. Item characteristic curves as estimated without knowledge of their mathematical form -- a confrontation of Birnbaum's logistical model. *Psychometrika*, 1970, 35,43-50.

[5] Lord, F.M. and M.R. Novick. *Statistical theories of mental test scores*. Addison-Wesley, 1968, Chapters 17-20.

[6] Rasch, G. *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedogogishe Institut, 1960.

[7] Samejima, F. A method of estimating item characteristic functions using the maximum likelihood estimate of ability. *Psychometrika*, 1977a, 42, 163-191.

[8] Samejima F. Estimation of the operating characteristics of item response categories I: Introduction to the Two-Parameter Beta Method. *Office of Naval Research*, Research Report 77-1, 1977b.

[9] Samejima, F. Estimation of the operating characteristics of item response categories II: Further development of the Two-Parameter Beta Method. *Office of Naval Research*, Research Report 78-1, 1978a.

[10] Samejima, F. Estimation of the operating characteristics of item response categories III: The Normal Approach Method and the Pearson System Method. *Office of Naval Research*, Research Report 78-2, 1978b.

REFERENCES (Continued)

[11] Samejima, F. Estimation of the operating characteristics of item response categories IV: Comparison of the different methods. Office of Naval Research, Research Report 78-3, 1978c.

[12] Samejima, F. Estimation of the operating characteristics of item response categories V: Weighted Sum Procedure in the Conditional P.D.F. Approach. Office of Naval Research, Research Report 78-4, 1978d.

[13] Samejima, F. Estimation of the operating characteristics of item response categories VI: Proportioned Sum Procedure in the Conditional P.D.F. Approach. Office of Naval Research, Research Report 78-5, 1978e.

[14] Samejima, F. Estimation of the operating characteristics of item response categories VII: Bivariate P.D.F. Approach with Normal Approach Method. Office of Naval Research, Research Report 78-6, 1978f.

[15] Samejima, F. A new family of models for the multiple-choice item. Office of Naval Research, Research Report 79-4, 1979.

[16] Samejima, F. Research on the multiple-choice test item in Japan: Toward the validation of mathematical models. Office of Naval Research, Research Report ONRT M3, 1980.

APPENDIX I

Bibliography of Selected Research Dealing with the Theoretical Development, or Applications of, the Three-Parameter Normal Ogive, or Logistic, Model

(1) Selection Process

A search was made of some 10 journals, and of the technical reports received by the principal investigator since 1977, for articles concerning the three-parameter normal ogive, or logistic, model. The selection procedure for these journals was somewhat arbitrary, based on sources cited in the present work, as well as the authors' knowledge of journals likely to contain such research. The names of those journals and the span of years, volumes, and numbers searched for each journal are given in Table A-1-1.

The articles, reports, and texts found are here organized into applied and theoretical research categories. The applied research categories are paper-and-pencil testing and computerized adaptive testing; the more theoretical works are grouped into areas of general review and development, considerations of the three-parameter approach, Bayesian estimation, and simulation studies. Each has been briefly summarized as it pertains to the three-parameter latent trait model; this was not an easy task.

TABLE A-1-1

Journal Titles and the Respective Span of Years, Volumes, and Numbers Searched for Three-Parameter Model Research Papers

Journal Title	Year, Volume, Number	
	Initial	Final
American Educational Research Journal	1975,12,1 -	1980,17,1
Applied Psychological Measurement	1977, 1,1 -	1980, 4,2
British Journal of Mathematical and Statistical Psychology	1974,27,1 -	1980,33,1
Educational and Psychological Measurement	1965,25,1 -	1980,40,1
Journal of Applied Psychology	1975,60,1 -	1980,65,2
Journal of Educational Measurement	1974,11,1 -	1980,17,1
Journal of Educational Statistics	1976, 1,1 -	1980, 5,1
Journal of Mathematical Psychology	1974,11,1 -	1980,21,2
Psychometrika	1965,30,1 -	1980,45,1
Review of Educational Research	1974,44,1 -	1980,50,1

(2) *Applied Research: Paper-and-Pencil Testing*

Lord, F. M. An empirical study of item-test regression. *Psychometrika*, 1965, 30, 373-376.

On a group of 103,275 examinees who each took a 90 item Verbal and 60 item Mathematical SAT, Lord examines the item-test regressions. He finds, among other things, that examinees with low observed scores may score significantly below chance levels when taking difficult items. Lord suggests that distractors may be working in these cases.

Lord, F. M. An analysis of the verbal scholastic aptitude test using Birnbaum's three parameter logistic model. *Educational and Psychological Measurement*, 1968, 28, 989-1020.

The item responses of 2,862 examinees who took the Verbal SAT were analyzed using the three parameter model. Efficiency of the model in terms of accuracy of estimation was found to be relatively high, and comparisons of information functions based on subtests were made.

Lord, F. M. Item characteristic curves estimated without knowledge of their mathematical form--A confrontation of Birnbaum's logistic model. *Psychometrika*, 1970, 35, 43-50.

Five items from the Verbal SAT, which were administered to 103,275 examinees, are taken as examples, and the author's method of of estimating the true test score distribution for the given set of observed test scores is applied for both success and failure groups. Based upon these two estimated true score distributions, the estimated item characteristic function is obtained as the ratio of the density for the success group over the total density, and the subsequent transformation of the true test score to the latent ability. When compared to three-parameter logistic curves generated for these items, a close fit between the two was found.

Lord, F. M. Power scores estimated by item characteristic curves. *Educational and Psychological Measurement*, 1973, 33, 219-224.

Using information from a re-administered 90-item Verbal SAT subtest that was originally mistimed for 21 examinees (they were originally tested under far too speeded conditions), Lord estimated ability levels of the examinees.

Lord, F. M. Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, 1974, 39, 247-264.

Using the three-parameter model, Lord analyzed 3 data sets (2,926 examinees on a 90 item verbal aptitude test; 994 examinees on a parallel form of that test; 2,946 examinees on an 85 item arithmetic reasoning test) with a formula-scoring method to maximize the available information.

Jensema, C. J. An application of latent trait mental test theory. *British Journal of Mathematical and Statistical Psychology*, 1974, 27, 29-48.

Item responses of 4,950 examinees to 94 mathematics questions in the Washington Pre-College Test battery were used to test several hypotheses. Results indicated that (1) Bayesian tailored testing does not lead to substantial reduction in item number in realistic situations; (2) parameter estimation might be accomplished graphically; (3) item banks should have wide difficulty ranges and have highly discriminating items.

Lord, F. M. The 'Ability' scale in item characteristic curve theory. *Psychometrika*, 1975, 40, 205-217.

Using 6 data sets (2,986 V-GRE, 2,862 V-SAT, 2,926 V-SAT, 2,946 M-SAT, 2,848 M-SAT, and 1,825 Vocabulary MAT at the sixth grade level), Lord found significant correlations over items between difficulty and discrimination parameters. He then proposed a transformation that would eliminate this problem.

Marco, G. L. Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*, 1977, 14, 139-160.

Using some simulation data in conjunction with 1,209 CLEP and 1,260 APP calculus test examinees, Marco applied the parameter estimation program LOGIST to the tasks of designing a multi-purpose test, evaluating a multi-level test, and equating a test on the basis of pretest statistics. Information functions are applied to the two initial problems.

Reckase, M. D. Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 1979, 4, 207-230.

Using 3,126 examinees who took verbal and quantitative subtests of the Missouri Scholastic Aptitude Test, and 1,000 simulated examinees with specific factor-structured hypothetical tests, Reckase found that (1) one- and three-parameter models estimate different factors when the test has independent factors, but estimate the first principal component when it is relatively large; (2) when independent factors exist, the three-parameter model estimates only one of them, ignoring the others, while the one parameter model seems to estimate the sum of the factors.

Trabin, T. E. and D. J. Weiss. The person response curve: Fit of individuals to item characteristic curve models. *Office of Naval Research*, Research Report 79-7, 1979.

Using the responses of 151 college students to 216 vocabulary test items, the authors were able to show that three-parameter logistic person response curves were good predictors of test response profiles for their data in more than 90% of the cases.

(3) Applied Research: Computerized Adaptive Testing

Koch, W. R. and M. D. Reckase. A live tailored testing comparison study of the one- and three- parameter logistic models. Office of Naval Research. Research Report 78-1, 1978.

Vocabulary test items were administered in an adaptive manner by computer to 128 college students. Comparison of relative efficiency curves, test-retest reliabilities, goodness of fit of the models, convergence rates, and criterion validity indicated the superiority of the three-parameter model over the one-parameter model.

Bejar, I. I. and D. J. Weiss. A construct validation of adaptive achievement testing. Office of Naval Research, Research Report 78-4, 1978.

Using two independent groups of 269 and 230 college students, the authors administered two stradaptive and two paper and pencil tests of achievement and vocabulary. Construct validities of the two modes of presentation were approximately equal, with the stradaptive tests scored by maximum likelihood estimation using the three-parameter logistic model requiring 25% to 31% fewer items.

Koch, W. R. and M. D. Reckase. Problems in application of latent trait models to tailored testing. Office of Naval Research, Research Report 79-1. 1979.

Giving a counterbalanced presentation of adaptive and paper-and-pencil achievement tests to 110 college students, the authors found that neither the one- nor the three- parameter logistic model scoring of the adaptive tests yielded satisfactory content validities. Further, the reliabilities of the adaptive tests were lower than the comparable paper and pencil test presentation of the multidimensional achievement test.

McKinley, R. L. and M. D. Reckase. A successful application of latent trait theory to tailored achievement testing. Office of Naval Research. Research Report 80-1, 1980.

In a replication of the above study corrected for improper item linking and selection procedures, the authors tested 88 college students with a multidimensional achievement test in an adaptive and a paper-and-pencil counterbalanced test-retest procedure. This time, they found that the one- and three- parameter tests had higher reliabilities over the same material than did a paper and pencil test. The three-parameter model yielded, reasonably enough, higher test information than did either the one parameter model adaptive test or the paper and pencil test. Neither adaptive procedure, however, yielded satisfactory content validity.

(4) Theoretical Research: General Review or Development

Baker, F. B. Advances in item analysis. Review of Educational Research, 1977, 47, 151-178.

Baker discusses and reviews research in latent trait estimation, including maximum likelihood approaches, for dichotomous and polychotomous scoring techniques, and describes developments in various models, including the Rasch and Three-Parameter cases.

Hambleton, R. K. and L. L. Cook. Latent trait models and their use in the analysis of educational test data. Journal of Educational Measurement, 1977, 14, 75-96.

The authors discuss the utility of various latent trait approaches to educational situations, with emphasis on logistic models, information functions, and parameter estimation procedures.

Lord, F. M. Practical applications of item characteristic curve theory. Journal of Educational Measurement, 1977, 14, 117-138.

Lord discusses procedures for designing, or redesigning, a test for latent trait use with the aid of information functions; item bias, two-stage and adaptive testing, and test equating are also discussed and referenced.

Urry, V. W. Tailored testing: A successful application of latent trait theory. Journal of Educational Measurement, 1977, 14, 181-196.

This paper discusses the necessary conditions for economical and efficient testing of uni- or multi-dimensional abilities with tailored testing, and describes possible applications to guidance and counseling, achievement testing, and computer-aided instruction.

Hambleton, R. K., Swaminathan, H., Cook, L. L., Fignor, D. R., and J. A. Gifford. Developments in latent trait theory: Models, technical issues, and applications. Review of Educational Research, 1978, 48, 467-510.

In a thorough review of the field, the authors describe the full range of models employed or developed, various parameter estimation procedures (including maximum likelihood and Bayesian), the use of information functions, and various applications of the models. They note in reference to the guessing parameter of the three-parameter model, that stability of the parameter can be found only when the examinee population is quite heterogeneous for the ability.

Warm, T. A. A primer of item reponse theory. *Department of Transportation*, Technical Report 941072, 1978.

This primer gives a general, expository presentation of latent trait theory and specific techniques of application, suggesting use of available computer programs where available and giving computation procedures elsewhere.

Lord, F. M. *Applications of item response theory to practical testing problems*. Hillsdale, New Jersey: Lawrence Erlbaum Assoc., 1980.

In this text, Lord presents a wide-ranging coverage of most topics involved in the adaptation of latent trait theory to real testing situations. He integrates and updates his past research, as well as that of most others in the field, and includes rationales and justifications for various applica-applications and techniques.

(5) Theoretical Research: Considerations of the Three-Parameter Approach

Samejima, F. A general model for free-response data. *Psychometrika Monograph*, 1972, No. 18.

Samejima integrates the dichotomous response level into her general model for free-response data, and shows that, since the three-parameter model has a lower asymptote greater than zero for the correct-answer category, the basic function cannot be strictly decreasing in ability, and, therefore, the unique maximum condition cannot be satisfied. This implies that there may not exist unique maximum likelihood estimates for some response patterns.

Samejima, F. A comment on Birnbaum's three-parameter logistic model in the latent trait theory. *Psychometrika*, 1973, 38, 221-233.

Based upon the fact that Birnbaum's three-parameter logistic model does not assure a unique maximum for the likelihood function for every response pattern, the critical value, $\underline{\theta}_g$, of ability θ for each item g is introduced. It is suggested that item g should not be used for estimating the examinee's ability below this point.

van der Linden, W. Forgetting, guessing, and mastery: The Macready and Dayton models revisited and compared with a latent trait approach. *Journal of Educational Statistics*, 1978, 3, 305-317.

Comparisons are drawn and distinctions made by the author in a comparison of Guttman-like mastery models with the three-parameter logistic model. The latent trait model is extended with the conception of mastery as a region on a latent variable.

(6) Theoretical Research: Bayesian Estimation

Jensema, C. J. The validity of Bayesian tailored testing. *Educational and Psychological Measurement*, 1974, 34, 757-766.

Using sets of 100 Monte-Carlo simulated examinees on four different item pools, Jensema found that the standard error of the estimate was a more accurate criterion for terminating item presentation than was number of items presented; the only exception to this generalization occurs when the item discrimination parameters roughly equal the guessing parameters.

McBride, J. R. Some properties of a Bayesian adaptive ability testing strategy. *Applied Psychological Measurement*, 1977, 1, 121-140.

In a series of 4 studies, with finite or infinite item pools and with normally distributed population or distribution-free context criteria, McBride examines the characteristics of Owen's Bayesian procedures for adaptive testing on simulated data. He finds a high correlation between the estimate of ability and the true ability, but also finds that the estimate is positively related to number of presented items, is biased, and varies in accuracy with respect to ability.

(7) Theoretical Research: Simulation Studies

Hambleton, R. K. and R. E. Traub. Information curves and efficiency of three logistic test models. *British Journal of Mathematical and Statistical Psychology*, 1971, 24, 273-281.

Using guessing-free simulation data, the authors show that the relative efficiency of the one-, two-, and three-parameter models, as defined as the ratio of information functions between two of them, is quite close if the range of item discrimination parameters is restricted, regardless of ability level. With guessing introduced, the three-parameter model is the much more efficient at lower ability levels.

Lord, F. M. Robbins-Monro procedures for tailored testing. *Educational and Psychological Measurement*, 1971, 31, 3-31.

Lord demonstrates the use of a shrinking-step-size procedure for selection of subsequent items in tailored testing, using a theoretical item pool with known item parameters for the three-parameter logistic model. He shows advantages of this procedure over fixed-step-width procedures, but indicates that item pool size increases exorbitantly in this approach if more than six or seven items are to be administered per examinee.

Lord, F. M. A theoretical study of two-stage testing. Psychometrika, 1971, 36, 227-242.

Using 200 different simulated two-stage testing designs, Lord finds some of them superior to adaptive testing, if guessing is not possible in the first stage of testing. Where guessing correctly occurs 20 percent of the time or more on the first stage, no two-stage procedure is as effective as tailored testing that can adapt up and down the ability spectrum.

Urry, V. W. Approximations to item parameters of mental test models and their uses. Educational and Psychological Measurement, 1974, 34, 253-269.

Urry proposes procedures for the initial approximation of the item parameters in the three-parameter models for use in item screening for adaptive testing and as first values for maximum likelihood estimation. The guessing parameter is approximated by visual inspection of the lower asymptote of the item-test regression curve. This parameter is then used to convert point-biserial item-test correlations to discrimination estimates and proportion correct to difficulty estimates.

Jensema, C. A simple technique for estimating latent trait mental test parameters. Educational and Psychological Measurement, 1976, 36, 705-715.

Following Urry's lead, Jensema describes straight-forward techniques to estimate three-parameter model parameters. In comparison with maximum likelihood estimates, based on 2,263 simulated items, the simple estimates correlated 0.86 for the discrimination parameter and 0.97 for the difficulty parameter. Accuracy of the technique increases as sample size and item number increase, and decreases as item discrimination increases.

Lord, F. M. Optimal number of choices per item-- A comparison of four approaches. Journal of Educational Measurement, 1977, 14, 33-38.

Under various [illegible] conditions, 90 items of the Verbal section of the S[illegible] manipulated such that their guessing parameters are eit[illegible] left as is, or changed to conform to 5-, 4-, 3-, or 2-choice items. Lord finds ability-specific results for these conditions. Decreasing the effective number of choices per item, while lengthening the test in proportion, increases the efficiency of the test for high-level examinees, who guess little if at all, but decreases it for low-ability examinees, who might be expected to guess more frequently.

Schmidt, F. L. The Urry method of approximating the item parameters of latent trait theory. Educational and Psychological Measurement, 1977, 37, 613-620.

Schmidt shows that Urry's (1974) method of parameter estimation underestimates the discrimination parameter, and overestimates the difficulty parameter. He also demonstrates a simple procedure to correct these biases by transforming the point-biserial correlation.

Levine, M. V. and D. B. Rubin. Measuring the appropriateness of multiple-choice test scores. Journal of Educational Statistics, 1979, 4, 269-290.

Using several indices of internal consistency and three-parameter model assumptions, Levine and Rubin explore simulated examinee item response patterns to SAT questions. They find that sufficient information does exist on the answer sheet alone to judge the appropriateness of the test for a given examinee or group of examinees.

Kingsbury, G. G. and D. J. Weiss. Relationships among achievement level estimates from three item characteristic curve scoring methods. Office of Naval Research, Research Report 79-3, 1979.

The authors compare maximum likelihood normal, maximum likelihood logistic, and Bayesian scoring methods using one-, two-, and three-parameter models for various data sets. The results indicate that for the three-parameter model, Bayesian estimation produces results divergent from the maximum likelihood techniques, but does not suffer from problems of non-convergence; convergence problems are maximal for maximum likelihood procedures with conventional tests of inappropriate difficulty.

Ree, M. J. Estimating item characteristic curves. Applied Psychological Measurement, 1979, 3, 371-385.

Ree uses 2,000 simulated subjects' responses to 80-item tests to judge the effectiveness of 4 different computer estimation procedures, where the distribution of ability is rectangular, random, or normal. OGIVIA was found to work best with the latter data set; LOGIST on the random distribution works most accurately, but costs substantially more in computer time.

APPENDIX II

Directions of Iowa Tests of Basic Skills for Each of the Eleven Subtests

DIRECTIONS FOR EACH OF THE ELEVEN SUBTESTS

(1) Subtest V

In each exercise, you are to decide which one of the four answers has most nearly the same meaning as the word in heavy type above them. Then, on the answer sheet, find the row of answer spaces numbered the same as the exercise you are working on. You are to fill in the answer space on the answer sheet that has the same number as the answer you picked. The sample exercise in the box at the right has already been marked correctly on the answer sheet.

Sample Exercise: **Close** the door
1) shut
2) hold
3) behind
4) open [answer: 1)]

(2) Subtest R

This test consists of several reading selections. After each selection there are some exercises. Read each selection quickly and then answer the exercises. Four answers are given for each exercise, but only one of these answers is right. You are to choose the one answer that you think is better than the others. Then, on the answer sheet, find the row of answer spaces numbered the same as the exercise. Fill in the answer space for the best answer. The sample exercise below shows you how to mark your answers on the answer sheet.

Sample Exercise: Every Sunday after dinner Pop gets a ball game on TV. The next thing we know he is snoring.

S1. What does Pop do on Sunday afternoon?
1) Works in the yard
2) Goes to church
3) Takes a nap
4) Plays ball [answer: 3)]

(3) Subtest L1

The exercises in this spelling test are like the samples shown at the right. Many of the exercises contain a mistake in spelling. Some do not have any mistakes at all. You are to look for mistakes in spelling. When you find a mistake, fill in the answer space on the answer sheet that has the same number as the word which is wrong. If there is no mistake in an exercise, fill in the fifth answer space. The sample exercises at the right show you what to do.

Sample Exercises:

S1. 1) our
2) mi
3) your
4) them
5) (No mistakes) [answer: 2)]

S2. 1) fill
2) keep
3) was
4) saw
5) (No mistakes) [answer: 5)]

(4) Subtest L2

This is a test on capitalization. It will show whether you know which words in a sentence should be capitalized. The exercises in the test are like the samples shown below. Many of the exercises contain mistakes in capitalization. Some do not have any mistakes at all. You are to look for mistakes in the test exercises. When you find a mistake, fill in the answer space on the answer sheet that has the same number as the line containing the mistake. If there is no mistake in an exercise, fill in the fourth answer space. The sample exercises below show what you are to do.

Sample Exercises:

S1. 1) Tom and jerry
2) picked up all the
3) trash from the picnic.
4) (No mistakes) [answer: 1)]

S2.
1) Sally said that
2) everyone should have
3) been more Careful.
4) (No mistakes) [answer: 3)]

S3.
1) Let's all help
2) to keep our streets
3) and sidewalks clean.
4) (No mistakes) [answer: 4)]

(5) Subtest L3

This is a test on punctuation. It will show how well you can use periods, commas, question marks, apostrophes, etc. The exercises in the test are like the samples shown below. Many of the exercises contain mistakes in punctuation. Some do not have any mistakes at all. You are to look for mistakes in the test exercises. When you find a mistake, fill in the answer space on the answer sheet that has the same number as the line containing the mistake. If there is no mistake in an exercise, fill in the fourth answer space. The sample exercises below show what you are to do.

Sample Exercises:

S1.
1) Our family tries
2) to practice
3) rules of safety
4) (No mistakes) [answer: 3)]

S2.
1) We all fasten
2) our seat belts
3) before, we leave.
4) (No mistakes) [answer: 3)]

S3.
1) We do our best
2) to make our home
3) a safe place to live.
4) (No mistakes) [answer: 4)]

(6) Subtest L4

This is a test on the use of words. It will show whether you know how to use words according to the standards of correctly written English. The exercises in the test are like the samples shown below. Many of the exercises contain mistakes in the use of words. Some do not have any mistakes at all. You are to look for mistakes in the test exercises. When you find a mistake, fill the answer space on the answer sheet that has the same number as the line containing the mistake. If there is no mistake in an exercise, fill in the fourth answer space. The sample exercises below show what you are to do.

Sample Exercises:

S1.
1) He showed us the way.
2) Are you afraid to try?
3) Me and him took turns.
4) (No mistakes) [answer: 3)]

S2.
1) Time went first.
2) The bird flew away.
3) Pat found a dollar.
4) (No mistakes) [answer: 4)]

(7) Subtest W1

This is a test of your ability to read maps. It contains several maps, with some exercises about each one. Four answers are given for each exercise, but only one of these answers is right. You are to choose the one answer that you think is better than the others. Then, on the answer sheet, find the row of answer spaces numbered the same as the exercise. Fill in the answer space for the best answer.

(8) Subtest W2

This is a test of your ability to read graphs and tables. After each graph or table there are several exercises. For each exercise, decide which answer is correct. Then mark the proper answer space on the answer sheet. Mark only one answer space for each exercise.

(9) Subtest W3

This is a test of study skills such as looking up words, alphabetizing, using an index, and locating information. Read the directions for each part carefully, and then mark your answers to the exercises on your answer sheet.

(10) Subtest M1

This is a test of how well you understand the number system and the terms and operations used in mathematics. Four answers are given for each exercise, but only one of these answers is right. You are to choose the one answer that you think is better than the others. Then, on the answer sheet, find the row of answer spaces numbered the same as the exercise. Fill in the answer space for the best answer.

(11) Subtest M2

This is a test of your skill in solving mathematics problems. The exercises in the test are like the samples shown at the right. After each exercise are three possible answers and a "Not given"--meaning that the correct answer is not given. Work each exercise and compare your answer with the three possible answers. If the correct answer is given, fill in the answer space on the answer sheet that has the same number as the right answer. If the correct answer is not given, fill in the fourth answer space. The sample exercises show you what to do.

Sample Exercises:

S1. Peg has 1 sister and 2 brothers.
How many brothers and sisters
does she have?
1) 2 3) 4
2) 3 4) (Not given) [answer: 2)]

S2. Ben had 5 butterflies in a jar.
He opened the jar and 4 flew
away. How many did he have left?
1) 5 3) 2
2) 4 4) (Not given) [answer: 4)]

APPENDIX III

Relative Frequencies of Examinees Who Left the Items Unanswered in Each Subtest, Arranged in the Order of Presentation of the Items

TABLE A-3-1

Relative Frequencies of Examinees Who Left the Items Unanswered in Each Subtest, Arranged in the Order of Presentation of the Items: Level 11.

Subtest V	0.0000	0.0008	0.0013	0.0013	0.0025	0.0008	0.0017	0.0021
	0.0017	0.0017	0.0021	0.0008	0.0008	0.0017	0.0038	0.0051
	0.0021	0.0030	0.0008	0.0000	0.0051	0.0030	0.0017	0.0025
	0.0034	0.0055	0.0072	0.0055	0.0118	0.0110	0.0165	0.0140
	0.0228	0.0271	0.0343	0.0381	0.0406	0.0444	0.0596	0.0626
	0.0660	0.0740	0.0863					
Subtest R	0.0000	0.0004	0.0025	0.0034	0.0013	0.0013	0.0042	0.0008
	0.0004	0.0000	0.0000	0.0004	0.0008	0.0008	0.0004	0.0000
	0.0042	0.0000	0.0017	0.0013	0.0008	0.0000	0.0013	0.0004
	0.0004	0.0000	0.0017	0.0013	0.0000	0.0025	0.0000	0.0008
	0.0021	0.0013	0.0000	0.0004	0.0008	0.0008	0.0034	0.0017
	0.0021	0.0030	0.0080	0.0034	0.0034	0.0051	0.0068	0.0106
	0.0068	0.0085	0.0089	0.0110	0.0118	0.0135	0.0131	0.0266
	0.0288	0.0300	0.0326	0.0326	0.0364	0.0402	0.0419	0.0453
	0.0486	0.0520	0.0584	0.0647	0.0698	0.0740	0.0808	0.0850
	0.0905	0.0931						
Subtest L1	0.0000	0.0000	0.0013	0.0004	0.0008	0.0008	0.0000	0.0004
	0.0008	0.0013	0.0008	0.0008	0.0013	0.0013	0.0013	0.0008
	0.0008	0.0017	0.0008	0.0008	0.0004	0.0021	0.0017	0.0030
	0.0038	0.0055	0.0076	0.0110	0.0102	0.0144	0.0199	0.0258
	0.0305	0.0360	0.0402	0.0486	0.0592	0.0706	0.0825	0.0931
	0.1003	0.1129	0.1273					
Subtest L2	0.0000	0.0008	0.0013	0.0008	0.0008	0.0008	0.0000	0.0004
	0.0013	0.0008	0.0017	0.0017	0.0004	0.0008	0.0008	0.0008
	0.0017	0.0013	0.0004	0.0000	0.0000	0.0030	0.0038	0.0042
	0.0047	0.0072	0.0076	0.0093	0.0097	0.0118	0.0152	0.0178
	0.0237	0.0250	0.0266	0.0368	0.0431	0.0499	0.0571	0.0635
Subtest L3	0.0000	0.0000	0.0030	0.0004	0.0013	0.0004	0.0004	0.0004
	0.0004	0.0008	0.0025	0.0030	0.0008	0.0004	0.0008	0.0017
	0.0008	0.0008	0.0000	0.0000	0.0008	0.0000	0.0017	0.0008
	0.0013	0.0034	0.0030	0.0051	0.0051	0.0055	0.0051	0.0068
	0.0085	0.0102	0.0085	0.0089	0.0127	0.0135	0.0144	0.0157

TABLE A-3-1 (Continued): Level 11.

Subtest L4	0.0000	0.0000	0.0030	0.0004	0.0030	0.0017	0.0013	0.0013
	0.0008	0.0013	0.0021	0.0017	0.0000	0.0008	0.0017	0.0021
	0.0000	0.0008	0.0013	0.0008	0.0004	0.0004	0.0004	0.0017
	0.0025	0.0038	0.0076	0.0076	0.0076	0.0080	0.0097	0.0114
Subtest W1	0.0000	0.0000	0.0008	0.0025	0.0030	0.0004	0.0013	0.0055
	0.0004	0.0008	0.0000	0.0008	0.0025	0.0034	0.0008	0.0004
	0.0059	0.0085	0.0013	0.0047	0.0038	0.0072	0.0089	0.0106
	0.0178	0.0250	0.0351	0.0461	0.0626	0.0863	0.1180	0.1404
	0.1574	0.1840	0.2098	0.2310				
Subtest W2	0.0004	0.0000	0.0004	0.0004	0.0004	0.0004	0.0008	0.0013
	0.0004	0.0030	0.0008	0.0017	0.0004	0.0013	0.0025	0.0030
	0.0047	0.0051	0.0068	0.0089	0.0182	0.0266	0.0338	0.0398
	0.0474	0.0563						
Subtest W3	0.0004	0.0017	0.0030	0.0025	0.0021	0.0034	0.0030	0.0008
	0.0021	0.0021	0.0017	0.0021	0.0055	0.0055	0.0030	0.0030
	0.0017	0.0008	0.0030	0.0034	0.0013	0.0013	0.0008	0.0021
	0.0025	0.0021	0.0025	0.0025	0.0025	0.0038	0.0089	0.0102
	0.0161	0.0157	0.0220	0.0296	0.0355	0.0410	0.0495	0.0541
	0.0694	0.0719	0.0774	0.0846	0.0897	0.0977	0.1062	0.1096
	0.1172	0.1282	0.1375	0.1421	0.1497	0.1548	0.1616	0.1667
Subtest M1	0.0004	0.0013	0.0004	0.0008	0.0055	0.0008	0.0055	0.0013
	0.0110	0.0004	0.0008	0.0017	0.0017	0.0017	0.0030	0.0013
	0.0025	0.0021	0.0025	0.0008	0.0030	0.0008	0.0042	0.0089
	0.0055	0.0021	0.0059	0.0038	0.0080	0.0157	0.0228	0.0152
	0.0216	0.0207	0.0258	0.0292	0.0385	0.0423	0.0465	0.0643
	0.0690	0.0783						
Subtest M2	0.0000	0.0008	0.0008	0.0025	0.0038	0.0055	0.0004	0.0013
	0.0072	0.0042	0.0008	0.0051	0.0008	0.0008	0.0038	0.0030
	0.0038	0.0165	0.0085	0.0152	0.0250	0.0266	0.0355	0.0499
	0.0575	0.0753	0.1003	0.1210	0.1426			

TABLE A-3-2

Relative Frequencies of Examinees Who Left the Items Unanswered in Each Subtest, Arranged in the Order of Presentation of the Items: Level 12.

Subtest V	0.0004	0.0000	0.0041	0.0004	0.0008	0.0017	0.0004	0.0008
	0.0025	0.0004	0.0008	0.0012	0.0021	0.0000	0.0029	0.0021
	0.0017	0.0008	0.0017	0.0008	0.0012	0.0017	0.0000	0.0017
	0.0004	0.0004	0.0050	0.0041	0.0054	0.0050	0.0050	0.0054
	0.0108	0.0099	0.0087	0.0112	0.0170	0.0174	0.0240	0.0278
	0.0323	0.0356	0.0431	0.0464	0.0481	0.0510		
Subtest R	0.0000	0.0004	0.0008	0.0017	0.0012	0.0012	0.0008	0.0029
	0.0008	0.0033	0.0004	0.0025	0.0004	0.0008	0.0000	0.0004
	0.0021	0.0033	0.0008	0.0120	0.0000	0.0004	0.0021	0.0017
	0.0029	0.0029	0.0012	0.0017	0.0021	0.0017	0.0004	0.0004
	0.0012	0.0021	0.0017	0.0037	0.0021	0.0000	0.0008	0.0033
	0.0004	0.0008	0.0029	0.0017	0.0029	0.0033	0.0054	0.0046
	0.0054	0.0116	0.0128	0.0157	0.0186	0.0170	0.0199	0.0236
	0.0240	0.0286	0.0269	0.0274	0.0369	0.0414	0.0435	0.0506
	0.0526	0.0555	0.0588	0.0613	0.0663	0.0680	0.0700	0.0738
	0.0796	0.0796	0.0821	0.0833				
Subtest L1	0.0000	0.0004	0.0012	0.0037	0.0012	0.0004	0.0000	0.0000
	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0008	0.0021
	0.0025	0.0004	0.0000	0.0008	0.0012	0.0012	0.0012	0.0012
	0.0029	0.0033	0.0058	0.0066	0.0075	0.0104	0.0128	0.0191
	0.0233	0.0282	0.0323	0.0344	0.0493	0.0576	0.0688	0.0796
	0.0932	0.1015	0.1111	0.1202	0.1281	0.1359		
Subtest L2	0.0000	0.0004	0.0008	0.0004	0.0004	0.0025	0.0000	0.0008
	0.0000	0.0012	0.0012	0.0012	0.0000	0.0004	0.0008	0.0012
	0.0017	0.0004	0.0004	0.0000	0.0000	0.0008	0.0012	0.0004
	0.0000	0.0008	0.0012	0.0033	0.0062	0.0066	0.0087	0.0104
	0.0133	0.0162	0.0224	0.0236	0.0269	0.0323	0.0361	0.0385
	0.0423	0.0439						
Subtest L3	0.0000	0.0000	0.0004	0.0000	0.0004	0.0017	0.0004	0.0000
	0.0008	0.0000	0.0021	0.0025	0.0004	0.0000	0.0008	0.0012
	0.0012	0.0017	0.0004	0.0004	0.0017	0.0017	0.0017	0.0021
	0.0008	0.0012	0.0017	0.0012	0.0012	0.0021	0.0021	0.0033
	0.0041	0.0041	0.0062	0.0058	0.0083	0.0099	0.0120	0.0128
	0.0120	0.0120						

TABLE A-3-2 (Continued): Level 12.

Subtest L4	0.0000	0.0000	0.0012	0.0000	0.0008	0.0000	0.0008	0.0029
	0.0004	0.0004	0.0017	0.0041	0.0004	0.0008	0.0017	0.0012
	0.0004	0.0004	0.0012	0.0012	0.0000	0.0004	0.0012	0.0025
	0.0017	0.0037	0.0033	0.0037	0.0017	0.0025	0.0037	0.0050
Subtest W1	0.0041	0.0025	0.0004	0.0041	0.0012	0.0021	0.0008	0.0004
	0.0012	0.0004	0.0017	0.0012	0.0004	0.0008	0.0025	0.0037
	0.0000	0.0017	0.0021	0.0012	0.0017	0.0095	0.0091	0.0124
	0.0199	0.0253	0.0265	0.0327	0.0365	0.0452	0.0506	0.0709
	0.0961	0.1169	0.1372	0.1625	0.1811	0.2060	0.2366	0.2482
Subtest W2	0.0004	0.0021	0.0008	0.0004	0.0004	0.0000	0.0012	0.0000
	0.0008	0.0000	0.0008	0.0012	0.0000	0.0004	0.0017	0.0004
	0.0017	0.0025	0.0046	0.0087	0.0153	0.0207	0.0249	0.0298
	0.0493	0.0634	0.0771	0.0866				
Subtest W3	0.0012	0.0000	0.0017	0.0025	0.0008	0.0008	0.0025	0.0008
	0.0004	0.0000	0.0004	0.0017	0.0017	0.0012	0.0004	0.0029
	0.0008	0.0012	0.0000	0.0008	0.0033	0.0021	0.0021	0.0008
	0.0004	0.0008	0.0000	0.0008	0.0012	0.0012	0.0017	0.0008
	0.0017	0.0025	0.0021	0.0029	0.0029	0.0033	0.0050	0.0046
	0.0050	0.0070	0.0099	0.0133	0.0145	0.0182	0.0178	0.0224
	0.0278	0.0315	0.0410	0.0410	0.0472	0.0522	0.0539	0.0597
	0.0655	0.0696	0.0750					
Subtest M1	0.0000	0.0008	0.0025	0.0004	0.0017	0.0012	0.0004	0.0008
	0.0025	0.0025	0.0025	0.0012	0.0021	0.0012	0.0008	0.0008
	0.0008	0.0000	0.0050	0.0008	0.0012	0.0012	0.0004	0.0033
	0.0012	0.0033	0.0037	0.0017	0.0075	0.0099	0.0066	0.0091
	0.0145	0.0133	0.0141	0.0178	0.0195	0.0220	0.0253	0.0286
	0.0381	0.0446	0.0448	0.0547	0.0555			
Subtest M2	0.0000	0.0000	0.0000	0.0037	0.0058	0.0008	0.0029	0.0025
	0.0012	0.0037	0.0008	0.0008	0.0033	0.0033	0.0041	0.0000
	0.0021	0.0041	0.0041	0.0054	0.0104	0.0104	0.0149	0.0220
	0.0274	0.0361	0.0464	0.0535	0.0700	0.0783	0.0974	

TABLE A-3-3

Relative Frequencies of Examinees Who Left the Items Unanswered in Each Subtest, Arranged in the Order of Presentation of the Items: Level 13.

Subtest V	0.0024	0.0016	0.0016	0.0016	0.0000	0.0000	0.0008	0.0004
	0.0016	0.0016	0.0016	0.0008	0.0004	0.0008	0.0045	0.0024
	0.0016	0.0000	0.0008	0.0020	0.0000	0.0008	0.0020	0.0012
	0.0008	0.0004	0.0016	0.0016	0.0024	0.0024	0.0032	0.0032
	0.0049	0.0049	0.0053	0.0049	0.0065	0.0069	0.0101	[illegible]
	0.0117	0.0122	0.0142	0.0158	0.0194	0.0207	0.0203	0.0198
Subtest R	0.0008	0.0000	0.0008	0.0008	0.0012	0.0016	0.0024	0.0020
	0.0016	0.0012	0.0008	0.0000	0.0020	0.0000	0.0004	0.0012
	0.0008	0.0016	0.0032	0.0117	0.0004	0.0004	0.0008	0.0012
	0.0008	0.0024	0.0012	0.0008	0.0028	0.0012	0.0000	0.0016
	0.0020	0.0041	0.0024	0.0004	0.0020	0.0020	0.0024	0.0008
	0.0008	0.0004	0.0004	0.0008	0.0020	0.0020	0.0016	0.0028
	0.0049	0.0041	0.0028	0.0041	0.0053	0.0061	0.0053	0.0061
	0.0065	0.0109	0.0142	0.0146	0.0122	0.0194	0.0215	0.0235
	0.0239	0.0267	0.0288	0.0304	0.0324	0.0401	0.0348	0.0381
	0.0417	0.0454	0.0458	0.0470	0.0474	0.0494		
Subtest L1	0.0000	0.0016	0.0012	0.0016	0.0012	0.0008	0.0004	0.0004
	0.0004	0.0036	0.0004	0.0008	0.0008	0.0008	0.0020	0.0012
	0.0008	0.0004	0.0008	0.0008	0.0024	0.0008	0.0020	0.0004
	0.0000	0.0036	0.0045	0.0057	0.0089	0.0097	0.0113	0.0130
	0.0166	0.0186	0.0223	0.0243	0.0296	0.0360	0.0413	0.0458
	0.0498	0.0531	0.0628	0.0705	0.0814	0.0871	0.0895	0.0944
Subtest L2	0.0004	0.0008	0.0008	0.0024	0.0012	0.0004	0.0024	0.0004
	0.0008	0.0000	0.0012	0.0020	0.0008	0.0000	0.0012	0.0020
	0.0016	0.0004	0.0000	0.0012	0.0004	0.0016	0.0004	0.0008
	0.0008	0.0024	0.0036	0.0024	0.0036	0.0041	0.0053	0.0057
	0.0093	0.0126	0.0122	0.0134	0.0162	0.0186	0.0203	0.0235
	0.0227	0.0235	0.0292					
Subtest L3	0.0004	0.0000	0.0008	0.0008	0.0016	0.0016	0.0000	0.0004
	0.0016	0.0000	0.0016	0.0032	0.0004	0.0008	0.0008	0.0016
	0.0008	0.0004	0.0008	0.0008	0.0016	0.0004	0.0004	0.0008
	0.0004	0.0004	0.0016	0.0016	0.0008	0.0016	0.0012	0.0008
	0.0016	0.0024	0.0041	0.0045	0.0036	0.0061	0.0081	0.0077
	0.0085	0.0089	0.0105					

TABLE A-3-3 (Continued): Level 13.

Subtest L4	0.0000	0.0000	0.0020	0.0008	0.0000	0.0000	0.0004	0.0012
	0.0012	0.0004	0.0004	0.0049	0.0004	0.0008	0.0012	0.0012
	0.0000	0.0004	0.0004	0.0016	0.0000	0.0000	0.0008	0.0008
	0.0012	0.0020	0.0016	0.0020	0.0012	0.0024	0.0036	0.0053
Subtest W1	0.0004	0.0000	0.0008	0.0020	0.0016	0.0012	0.0004	0.0008
	0.0032	0.0016	0.0028	0.0049	0.0008	0.0008	0.0032	0.0036
	0.0028	0.0024	0.0016	0.0004	0.0012	0.0041	0.0036	0.0045
	0.0105	0.0150	0.0166	0.0166	0.0267	0.0271	0.0425	0.0486
	0.0632	0.0697	0.0765	0.0851	0.1000	0.1158	0.1280	0.1393
	0.1507							
Subtest W2	0.0000	0.0000	0.0008	0.0004	0.0000	0.0004	0.0024	0.0008
	0.0004	0.0024	0.0008	0.0012	0.0008	0.0016	0.0016	0.0016
	0.0016	0.0020	0.0024	0.0049	0.0081	0.0101	0.0174	0.0219
	0.0288	0.0348	0.0409	0.0441				
Subtest W3	0.0008	0.0004	0.0020	0.0012	0.0024	0.0012	0.0024	0.0008
	0.0004	0.0004	0.0004	0.0016	0.0020	0.0000	0.0024	0.0004
	0.0004	0.0004	0.0004	0.0016	0.0012	0.0024	0.0000	0.0008
	0.0004	0.0016	0.0012	0.0008	0.0020	0.0032	0.0045	0.0028
	0.0032	0.0020	0.0004	0.0012	0.0028	0.0041	0.0049	0.0061
	0.0057	0.0097	0.0130	0.0142	0.0166	0.0166	0.0231	0.0227
	0.0267	0.0324	0.0369	0.0389	0.0425	0.0478	0.0490	0.0535
	0.0571	0.0563	0.0624					
Subtest M1	0.0012	0.0004	0.0036	0.0016	0.0012	0.0024	0.0000	0.0036
	0.0004	0.0008	0.0016	0.0028	0.0004	0.0024	0.0028	0.0012
	0.0004	0.0024	0.0004	0.0020	0.0028	0.0012	0.0024	0.0012
	0.0008	0.0004	0.0053	0.0004	0.0004	0.0016	0.0024	0.0024
	0.0036	0.0041	0.0105	0.0085	0.0089	0.0093	0.0122	0.0150
	0.0211	0.0243	0.0296	0.0328	0.0369	0.0429	0.0429	0.0498
Subtest M2	0.0000	0.0004	0.0012	0.0004	0.0012	0.0057	0.0008	0.0012
	0.0012	0.0024	0.0008	0.0004	0.0008	0.0028	0.0053	0.0004
	0.0032	0.0045	0.0049	0.0065	0.0101	0.0065	0.0077	0.0113
	0.0154	0.0198	0.0316	0.0425	0.0531	0.0648	0.0680	0.0761

APPENDIX IV

Frequency Distribution of Items for Each of the Eleven Subtests with Respect to the Probability Resultant from the Chi-Square Test of the Goodness of Fit, and the Percentage Correct Response for Each of Those Items Whose Resultant Probability Is 0.001 or Greater

TABLE A-4-1

Frequency Distribution of Items for Each of the Eleven Subtests with Respect to the Probability Resultant from the Chi-Square Test of the Goodness of Fit. The Uniform Distribution Is Assumed for the Theoretical Frequency Distribution for the Incorrect Alternatives. Number of Degrees of Freedom for Each Item Is 2, Except for the Items of Subtest L1 for Which the Number of Degrees of Freedom Is 3. (Original Data)

Level 11

Subtest	Probability										Total
	.0000-.0005	.0005-.0015	.0015-.0055	.0055-.0105	.0105-.0505	.0505-.1005	.1005-.2005	.2005-.4005	.4005-.8005	.3005-1.0000	
V	42					1					43
R	68	1	2			2				1	74
L1	43										43
L2	40										40
L3	40										40
L4	32										32
W1	31		1		1		1		1	1	36
W2	24	1			1						26
W3	53						2		1		56
M1	36	1					1	1	2	1	42
M2	29										29
Total	438	3	3	0	2	3	4	1	4	3	461

Level 12

Subtest	Probability										Total
	.0000-.0005	.0005-.0015	.0015-.0055	.0055-.0105	.0105-.0505	.0505-.1005	.1005-.2005	.2005-.4005	.4005-.8005	.8005-1.0000	
V	42	1		1	1			1			46
R	71	2		1	1				1		76
L1	46										46
L2	42										42
L3	42										42
L4	32										32
W1	37	1			1			1			40
W2	24	1	2					1			28
W3	57	1						1			59
M 1	42				1		1		1		45
M2	30			1							31
Total	465	6	2	3	4	0	1	4	2	0	487

Level 13

Subtest	Probability										Total
	.0000-.0005	.0005-.0015	.0015-.0055	.0055-.0105	.0105-.0505	.0505-.1005	.1005-.2005	.2005-.4005	.4005-.8005	.8005-1.0000	
[illegible]	45					1	1			1	48
R	71	1	1		2	1		1	1		78
L1	48										48
L2	43										43
L3	43										43
L4	32										32
W1	40					1					41
W2	25	1				1				1	28
W3	57					1	1				59
M1	44	1		1			1	1			48
M2	31				1						32
Total	479	3	1	1	3	5	3	2	1	2	500

TABLE A-4-2

Percentage Correct Response for Each of the Items Whose Probability Obtained by the Chi-Square Test of the Goodness of Fit Is 0.001 or Greater. (Original Data)

Level 11

	Subtest	Item Number	Probability	Percentage Correct
1	V	61	0.061	35.9
2	R	26	0.001**	93.1
3	R	42	0.081	92.9
4	R	73	0.002**	64.8
5	R	78	0.889	79.2
6	R	88	0.004**	58.2
7	R	47	0.063	81.1
8	W1	13	0.005**	93.7
9	W1	20	0.036*	90.7
10	W1	21	0.824	83.9
11	W1	25	0.514	57.9
12	W1	29	0.147	52.1
13	W2	29	0.012*	85.7
14	W2	41	0.001**	63.7
15	W3	39	0.565	63.7
16	W3	40	0.111	76.8
17	W3	53	0.142	63.3
18	M1	31	0.445	91.3
19	M1	37	0.121	53.2
20	M1	38	0.669	80.3
21	M1	45	0.246	68.0
22	M1	63	0.001**	29.9
23	M1	68	0.815	71.7

TABLE A-4-2 (Continued): Level 12.

	Subtest	Item Number	Probability	Percentage Correct
1	V	49	0.009**	65.4
2	V	52	0.001**	57.0
3	V	57	0.048*	63.3
4	V	79	0.209	46.6
5	R	78	0.584	89.6
6	R	88	0.009**	73.9
7	R	105	0.001**	63.5
8	R	127	0.001**	53.8
9	R	131	0.018*	54.4
10	W1	28	0.001**	69.4
11	W1	29	0.289	73.8
12	W1	53	0.022*	46.0
13	W2	44	0.261	85.2
14	W2	45	0.001**	72.2
15	W2	48	0.005**	60.5
16	W2	59	0.004**	55.9
17	W3	50	0.001**	75.8
18	W3	53	0.383	79.1
19	M1	68	0.016*	87.4
20	M1	74	0.171	79.3
21	M1	95	0.660	29.4
22	M2	59	0.007**	47.2

TABLE A-4-2 (Continued): Level 13.

	Subtest	Item Number	Probability	Percentage Correct
1	V	61	0.122	57.6
2	V	79	0.063	56.6
3	V	93	0.851	49.9
4	R	88	0.001**	80.1
5	R	92	0.032*	86.9
6	R	103	0.002**	71.1
7	R	131	0.098	63.2
8	R	140	0.403	33.8
9	R	142	0.021*	38.0
10	R	155	0.249	35.7
11	W1	78	0.066	35.7
12	W2	44	0.915	88.1
13	W2	45	0.001**	79.7
14	W2	48	0.096	66.1
15	W3	70	0.089	85.3
16	W3	83	0.186	50.5
17	M1	74	0.010*	87.2
18	M1	88	0.001**	73.6
19	M1	105	0.134	71.9
20	M1	107	0.280	40.6
21	M2	75	0.050	29.4

DISTRIBUTION LIST

Navy

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Breaux
Code N-711
NAVTRAEQUIPCEN
Orlando, FL 32813

1 COMNAVMILPERSCOM (N-6C)
Dept. of Navy
Washington, DC 20370

1 Dr. Larry Dean, LT, MSC, USN
Psychology Department
Naval Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 06340

1 Dr. Richard Elster
Department of Administrative Sciences
Naval Postgraduate School
Monterey, CA 93940

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Patrick R. Harrison
Psychology Course Director
LEADERSHIP & LAW DEPT. (7b)
DIV. OF PROFESSIONAL DEVELOPMMENT
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

Navy

1 Dr. William L. Maloy
Principal Civilian Advisor for
Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 Dr. Kneale Marshall
Scientific Advisor to DCNO(MPT)
OP01T
Washington DC 20370

1 CAPT Richard L. Martin, USN
Prospective Commanding Officer
USS Carl Vinson (CVN-70)
Newport News Shipbuilding and Drydock Co
Newport News, VA 23607

1 Dr. James McBride
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. George Moeller
Head, Human Factors Dept.
Naval Submarine Medical Research Lab
Groton, CN 06340

1 Library
Naval Health Research Center
P. O. Box 85122
San Diego, CA 92138

1 Naval Medical R&D Command
Code 44
National Naval Medical Center
Betheada, MD 20014

1 Dr. Ted M. I. Yellen
Technical Information Office, Code 201
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Library, Code P201L
Navy Personnel R&D Center
San Diego, CA 92152

1 Technical Director
Navy Personnel R&D Center
San Diego, CA 92152

Navy

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 Psychologist
ONR Branch Office
Bldg 114, Section D
666 Summer Street
Boston, MA 02210

1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

1 Office of Naval Research
Code 437
800 N. Quincy SStreet
Arlington, VA 22217

5 Personnel & Training Research Programs
(Code 458)
Office of Naval Research
Arlington, VA 22217

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Office of the Chief of Naval Operations
Research, Development, and Studies Branc
(OP-102)
Washington, DC 20350

1 Captain Donald F. Parker, USN
Commanding Officer
Navy Personnel R&D Center
San Diego, CA 92152

1 LT Frank C. Petho, MSC, USN (Ph.D)
Code L51
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508

Navy

1 Director, Research & Analysis Division
Plans and Policy Department
Navy Recruiting Command
4015 Wilson Boulevard
Arlington, VA 22203

1 Dr. Bernard Rimland (03B)
Navy Personnel R&D Center
San Diego, CA 92152

1 Mr. Arnold Rubenstein
Naval Personnel Support Technology
Naval Material Command (08T244)
Room 1044, Crystal Plaza #5
2221 Jefferson Davis Highway
Arlington, VA 20360

1 Dr. Worth Scanland
Chief of Naval Education and Training
Code N-5
NAS, Pensacola, FL 32508

1 Dr. Robert G. Smith
Office of Chief of Naval Operations
OP-987H
Washington, DC 20350

1 Dr. Alfred F. Smode
Training Analysis & Evaluation Group
(TAEG)
Dept. of the Navy
Orlando, FL 32813

1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Ronald Weitzman
Code 54 WZ
Department of Administrative Sciences
U. S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Wisher
Code 309
Navy Personnel R&D Center
San Diego, CA 92152

Navy

1 DR. MARTIN F. WISKOFF
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Mr John H. Wolfe
Code P310
U. S. Navy Personnel Research and
Development Center
San Diego, CA 92152

Army

1 Technical Director
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 HQ USAREUE & 7th Army
ODCSOPS
USAAREUE Director of GED
APO New York 09403

1 Col Gary W. Bloedorn
US Army TRADOC Systems Analysis Activity
Attn: ATAA-TH
WSMR, NM 88002

1 DR. RALPH DUSEK
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Myron Fischl
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Milton S. Katz
Training Technical Area
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Harold F. O'Neil, Jr.
Attn: PERI-OK
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 LTC Michael Plummer
Chief, Leadership & Organizational
Effectiveness Division
Office of the Deputy Chief of Staff
for Personnel
Dept. of the Army
Pentagon, Washington DC 20301

Army

1 DR. JAMES L. RANEY
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Mr. Robert Ross
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Robert Sasmor
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Commandant
US Army Institute of Administration
Attn: Dr. Sherrill
FT Benjamin Harrison, IN 46256

1 Dr. Frederick Steinheiser
U. S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

1 Air Force Human Resources Lab
AFHRL/MPD
Brooks AFB, TX 78235

1 U.S. Air Force Office of Scientific
Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, DC 20332

1 Air University Library
AUL/LSE 76/443
Maxwell AFB, AL 36112

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 Research and Measurment Division
Research Branch, AFMPC/MPCYPR
Randolph AFB, TX 78148

1 Dr. Malcolm Ree
AFHRL/MP
Brooks AFB, TX 78235

1 Dr. Marty Rockway (AFHRL/TT)
Lowry AFB
Colorado 80230

1 Dr. Frank Schufletowski
U.S. Air Force
ATC/XPTD
Randolph AFB, TX 78148

1 Jack A. Thorpe, Maj., USAF
Naval War College
Providence, RI 02846

1 Dr. Joe Ward, Jr.
AFHRL/MPMD
Brooks AFB, TX 78235

Marines

1 Dr. H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134

1 Major Howard Langdon
Headquarters, Marine Corps
OTTI 31
Arlington Annex
Columbia Pike at Arlington Ridge Rd.
Arlington, VA 20380

1 Director, Office of Manpower Utilization
HQ, Marine Corps (MPU)
BCB, Bldg. 2009
Quantico, VA 22134

1 Headquarters, U. S. Marine Corps
Code MPI-20
Washington, DC 20380

1 Special Assistant for Marine
Corps Matters
Code 100M
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

1 Major Michael L. Patrow, USMC
Headquarters, Marine Corps
(Code MPI-20)
Washington, DC 20380

1 DR. A.L. SLAFKOSKY
SCIENTIFIC ADVISOR (CODE RD-1)
HQ, U.S. MARINE CORPS
WASHINGTON, DC 20380

CoastGuard

1 Chief, Psychological Reserch Branch
U. S. Coast Guard (G-P-1/2/TP42)
Washington, DC 20593

1 Mr. Thomas A. Warm
U. S. Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

Other DoD

12 Defense Documentation Center
Cameron Station, Bldg. 5
Alexandria, VA 22314
Attn: TC

1 Dr. Dexter Fletcher
ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BLVD.
ARLINGTON, VA 22209

1 Dr. William Graham
Testing Directorate
MEPCOM/MEPCT-P
Ft. Sheridan, IL 60037

1 Director, Research and Data
OASD(MRA&L)
3B919, The Pentagon
Washington, DC 20301

1 Military Assistant for Training and
Personnel Technology
Office of the Under Secretary of Defense
for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 MAJOR Wayne Sellman, USAF
Office of the Assistant Secretary
of Defense (MRA&L)
3B930 The Pentagon
Washington, DC 20301

Civil Govt

1 Dr. Susan Chipman
Learning and Development
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Lorraine D. Eyde
Personnel R&D Center
Office of Personnel Management of USA
1900 EStreet NW
Washington, D.C. 20415

1 Dr. Jerry Lehnus
REGIONAL PSYCHOLOGIST
U.S. Office of Personnel Management
230 S. DEARBORN STREET
CHICAGO, IL 60604

1 Dr. Andrew R. Molnar
Science Education Dev.
and Research
National Science Foundation
Washington, DC 20550

1 Personnel R&D Center
Office of Personnel Managment
1900 E Street NW
Washington, DC 20415

1 Dr. H. Wallace Sinaiko
Program Director
Manpower Research and Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

1 Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E Street NW
Washington, DC 20415

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Non Govt

1 Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Dr. Jackson Beatty
Department of Psychology
University of California
Los Angeles, CA 90024

1 Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

1 Dr. John Bergan
School of Education
University of Arizona
Tuscon AZ 85721

1 Dr. Werner Birke
DezWPs im Streitkraefteamt
Postfach 20 50 03
D-5300 Bonn 2
WEST GERMANY

1 Dr. R. Darrel Bock
Department of Education
University of Chicago
Chicago, IL 60637

1 Dr. Nicholas A. Bond
Dept. of Psychology
Sacramento State College
600 Jay Street
Sacramento, CA 95819

1 Dr. Robert Brennan
American College Testing Programs
P. O. Box 168
Iowa City, IA 52240

Non Govt

1 Dr. C. VICTOR BUNDERSON
WICAT INC.
UNIVERSITY PLAZA, SUITE 10
1160 SO. STATE ST.
OREM, UT 84057

1 Dr. Anthony Cancelli
School of Education
University of Arizona
Tuscon, AZ 85721

1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Charles Myers Library
Livingstone House
Livingstone Road
Stratford
London E15 2LJ
ENGLAND

1 Dr. Kenneth E. Clark
College of Arts & Sciences
University of Rochester
River Campus Station
Rochester, NY 14627

1 Dr. Norman Cliff
Dept. of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

1 Dr. William E. Coffman
Director, Iowa Testing Programs
334 Lindquist Center
University of Iowa
Iowa City, IA 52242

1 Dr. Meredith P. Crawford
American Psychological Association
1200 17th Street, N.W.
Washington, DC 20036

Non Govt

1 Dr. Hans Crombag
Education Research Center
University of Leyden
Boerhaavelaan 2
2334 EN Leyden
The NETHERLANDS

1 Director
Behavioural Sciences Division
Defence & Civil Institute of
Environmental Medicine
Post Office Box 2000
Downsview, Ontario M3M 3B9
CANADA

1 LCOL J. C. Eggenberger
DIRECTORATE OF PERSONNEL APPLIED RESEARC
NATIONAL DEFENCE HQ
101 COLONEL BY DRIVE
OTTAWA, CANADA K1A 0K2

1 ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

1 Dr. Leonard Feldt
Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Dr. Richard L. Ferguson
The American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Univ. Prof. Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

1 Professor Donald Fitzgerald
University of New England
Armidale, New South Wales 2351
AUSTRALIA

Non Govt

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 DR. ROBERT GLASER
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 DR. JAMES G. GREENO
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ron Hambleton
School of Education
University of Massechusetts
Amherst, MA 01002

1 Dr. Chester Harris
School of Education
University of California
Santa Barbara, CA 9310

1 Dr. Frederick Hayes-Roth
The Rand Corporation
1700 Main Street
Santa Monica, CA 90406

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Library
HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921

1 Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

Non Govt

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

1 Dr. Huynh Huynh
College of Education
University of South Carolina
Columbia, SC 29208

1 Dr. Douglas H. Jones
Rm T-255
Educational Testing Service
Princeton, NJ 08450

3 Journal Supplement Abstract Service
American Psychological Association
1200 17th Street N.W.
Washington, DC 20036

1 Professor John A. Keats
University of Newcastle
AUSTRALIA 2305

1 Dr. Stephen Kosslyn
Harvard University
Department of Psychology
33 Kirkland Street
Cambridge, MA 02138

1 Mr. Marlin Kroger
1117 Via Goleta
Palos Verdes Estates, CA 90274

1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
Pittsburgh, PA 15260

1 Dr. Michael Levine
210 Education Building
University of Illinois
Champaign, IL 61820

Non Govt

1 Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat
Groningen
NETHERLANDS

1 Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. James Lumsden
Department of Psychology
University of Western Australia
Nedlands W.A. 6009
AUSTRALIA

1 Dr. Gary Marco
Educational Testing Service
Princeton, NJ 08450

1 Dr. Scott Maxwell
Department of Psychology
University of Houston
Houston, TX 77004

1 Dr. Samuel T. Mayo
Loyola University of Chicago
820 North Michigan Avenue
Chicago, IL 60611

1 Professor Jason Millman
Department of Education
Stone Hall
Cornell University
Ithaca, NY 14853

1 Dr. Melvin R. Novick
356 Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

Non Govt

1 Dr. Jesse Orlansky
Institute for Defense Analyses
400 Army Navy Drive
Arlington, VA 22202

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207

1 DR. DIANE M. RAMSEY-KLEE
R-K RESEARCH & SYSTEM DESIGN
3947 RIDGEMONT DRIVE
MALIBU, CA 90265

1 MINRAT M. L. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 1328
D-53 BONN 1, GERMANY

1 Dr. Mark D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
4 Hill Hall
Columbia, MO 65211

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Leonard L. Rosenbaum, Chairman
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Lawrence Rudner
403 Elm Avenue
Takoma Park, MD 20012

1 Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

Non Govt

1 DR. WALTER SCHNEIDER
DEPT. OF PSYCHOLOGY
UNIVERSITY OF ILLINOIS
CHAMPAIGN, IL 61820

1 DR. ROBERT J. SEIDEL
INSTRUCTIONAL TECHNOLOGY GROUP
HUMRRO
300 N. WASHINGTON ST.
ALEXANDRIA, VA 22314

1 Dr. Kazuo Shigemasu
University of Tohoku
Department of Educational Psychology
Kawauchi, Sendai 980
JAPAN

1 Dr. Edwin Shirkey
Department of Psychology
University of Central Florida
Orlando, FL 32816

1 Dr. Robert Smith
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903

1 Dr. Richard Snow
School of Education
Stanford University
Stanford, CA 94305

1 Dr. Kathryn T. Spoehr
Department of Psychology
Brown University
Providence, RI 02912

1 Dr. Robert Sternberg
Dept. of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 Dr. David Stone
ED 236
SUNY, Albany
Albany, NY 12222

Non Govt

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

1 Dr. Hariharan Swaminathan
Laboratory of Psychometric and Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

1 Dr. Brad Sympson
Psychometric Research Group
Educational Testing Service
Princeton, NJ 08541

1 Dr. Kikumi Tatsuoka
Computer Based Education Research Laboratory
252 Engineering Research Laboratory
University of Illinois
Urbana, IL 61801

1 Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

1 Dr. Douglas Towne
Univ. of So. California
Behavioral Technology Labs
1845 S. Elena Ave.
Redondo Beach, CA 90277

1 Dr. J. Uhlaner
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

1 Dr. Howard Wainer
Bureau of Social SCience Research
1990 M Street, N. W.
Washington, DC 20036

Non Govt

1 Dr. Phyllis Weaver
Graduate School of Education
Harvard University
200 Larsen Hall, Appian Way
Cambridge, MA 02138

1 Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

1 DR. SUSAN E. WHITELY
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
LAWRENCE, KANSAS 66044

1 Dr. Wolfgang Wildgrube
Streitkraefteamt
Box 20 50 03
D-5300 Bonn 2
WEST GERMANY

1 Dr. J. Arthur Woodward
Department of Psychology
University of California
Los Angeles, CA 90024

Navy

1 Mr. Donald Calder
Office of Naval Research
325 Hinman Research Building
Atlanta, GA 30332

Army

1 Dr. Randall M. Chambers
U.S. Army Research Institute
for the Behavioral & Social
Sciences
Fort Sill Field Unit
P.O. Box 3066
Fort Sill, OK 73503

Non Govt

1 Dr. Bert F. Green
Department of Psychology
The John's Hopkins University
Charles at 34th Street
Baltimore, MD 21218

1 Dr. Ron Hambleton
School of Education
University of Massachusetts
Amherst, MA 01002

1 Dr. William W. Turnbull
Educational Testing Service
Princeton, NJ 08540

1 Dr. Isaac I. Bejar
Department of Psychology
Elliot Hall
75 East River Road
Minneapolis, MN 55455

1 Dr. George Woods
1106 Newport Ave.
Victoria, B. C.
V8S 5E4 Canada

1 Dr. Lowell Schipper
Department of Psychology
Bowling Green State University
Bowling Green, OH 43403

Non Govt

1 Dr. F. Mengal
Faculte' de Psychologie
et des Sciences de l'Education
Universite' de Geneve
3 fl. de l'Universite
1201 Geneva SWITZERLAND

1 Dr. Wim J. van der Linden
Vakgroep Onderwisjskunde
Postbus 217
7500 EA Enschede
The Netherlands

1 Dr. Lutz Hornke
University Duesseldorf
Erz. Wiss.
D-4000 Duesseldorf
WEST GERMANY

1 Dr. Wolfgang Buchtala
8346 Simbach Inn
Postfach 1306
Industriestrasse 1
WEST GERMANY

1 Mr. Phillip S. Livingston
1500 Sparkman #32C
Huntsville, AL 35805

1 Dr. Sukeyori Shiba
Faculty of Education
University of Tokyo
Hongo, Bumkyoku
Tokyo, Japan 113

1 Mr. Yukihiro Noguchi
Faculty of Education
University of Tokyo
Hongo, Bumkyoku
Tokyo, Japan 113

1 Dr. Takahiro Sato (Representative)
Application Research Laboratory
Central Research Laboratories
Nippon Electric Co., Ltd.,
4-1-1 Miyazaki, Takatsu-ku
Kawasaki 213, Japan